W I M M E R &

D O M I N I C K

Mass Media Research

An Introduction

Here it is! The first book that gives you the specific information you need—as a student or professional in the mass media—to understand, analyze, and apply the data produced by mass media research. Inside this book you'll find:

mass media research

Applications

the specific concepts and techniques used in the print and electronic media, in advertising and public relations

mass media research

Procedures

with emphasis on sampling, plus basic statistical theory and practice, written for nonstatisticians

mass media research

Approaches

laboratory, survey, and field research methods and applications, including content analysis

Mass Media Research: An Introduction

by Roger D. Wimmer & Joseph R. Dominick

82 W 5300
0-534-01228-0

Mass Media Research

An Introduction

Roger D. Wimmer
Cox Communications, Inc.
Joseph R. Dominick
University of Georgia

Wadsworth Publishing Company
Belmont, California
A Division of Wadsworth, Inc.

For Chris and Joan

Senior Editor: Rebecca Hayden
Production Editor: Sally Schuman
Designer: MaryEllen Podgorski
Copy Editor: William Reynolds
Technical Illustrator: Harry Spitzer
Signing Representatives: Tom Orsi and
Mark Francisco

Printed in the United States of America

1 2 3 4 5 6 7 8 9 10—87 86 85 84 83

Library of Congress Cataloging in Publication Data

Wimmer, Roger D.
Mass media research.

Includes bibliographies and index.
1. Mass media—Methodology. 2.Mass media—Research. I. Dominick, Joseph R. II. Title.
P91.W46 1983 001.51'01'8 82-10920
ISBN 0-534-01228-0

Mass Media Research

Wadsworth Series in Mass Communication

Rebecca Hayden, Senior Editor

General

MEDIAMERICA: Form, Content, and Consequence of Mass Communication, 2d, by Edward Jay Whetmore

THE INTERPLAY OF INFLUENCE: Mass Media and Their Publics in News, Advertising, Politics by Kathleen Hall Jamieson and Karlyn Kohrs Campbell

MASS COMMUNICATION AND EVERYDAY LIFE: A Perspective on Theory and Effects by Dennis K. Davis and Stanley J. Baran

MASS MEDIA RESEARCH: An Introduction by Roger D. Wimmer and Joseph R. Dominick

THE INTERNSHIP EXPERIENCE by Lynne Schafer Gross

Telecommunications

BROADCASTING AND THE PUBLIC, 2d, by Harrison B. Summers, Robert E. Summers, and John H. Pennybacker

STAY TUNED: A Concise History of American Broadcasting by Christopher H. Sterling and John M. Kittross

BROADCAST PROGRAMMING: Strategies for Winning Television and Radio Audiences by Susan Tyler Eastman, Sydney W. Head, and Lewis Klein

ADVERTISING IN THE BROADCAST MEDIA by Elizabeth J. Heighton and Don R. Cunningham

STRATEGIES IN BROADCAST AND CABLE PROMOTION by Susan Tyler Eastman and Robert A. Klein

MODERN RADIO STATION PRACTICES, 2d, by Joseph S. Johnson and Kenneth K. Jones

THE MAGIC MEDIUM: An Introduction to Radio in America by Edward Jay Whetmore

AUDIO IN MEDIA by Stanley R. Alten

TELEVISION PRODUCTION HANDBOOK, 3d, by Herbert Zettl

TELEVISION PRODUCTION WORKBOOK, 3d, by Herbert Zettl

SIGHT-SOUND-MOTION: Applied Media Aesthetics by Herbert Zettl

Journalism

REPORTING PROCESSES AND PRACTICES: Newswriting for Today's Readers by Everette E. Dennis and Arnold H. Ismach

THE NEWSPAPER: An Introduction to Newswriting and Reporting by Ronald P. Lovell

EXCELLENCE IN COLLEGE JOURNALISM by Wayne Overbeck and Thomas M. Pasqua

NEWS EDITING IN THE '80s: Text and Exercises by William L. Rivers

REPORTING PUBLIC AFFAIRS: Problems and Solutions by Ronald P. Lovell

FREE-LANCER AND STAFF WRITER: Newspaper Features and Magazine Articles, 3d, by William L. Rivers and Shelley Smolkin

MAGAZINE EDITING IN THE '80s: Text and Exercises by William L. Rivers

NEWSWRITING FOR THE ELECTRONIC MEDIA: Principles, Examples, Applications by Daniel E. Garvey and William L. Rivers

THIS IS PR: The Realities of Public Relations, 2d, by Doug Newsom and Alan Scott

WRITING IN PUBLIC RELATIONS PRACTICE: Form and Style by Doug Newsom and Tom Siegfried

CREATIVE STRATEGY IN ADVERTISING by A. Jerome Jewler

PICTURES ON A PAGE: Photojournalism and Picture Editing by Harold Evans

PREFACE

Research in the mass media has expanded at a rapid rate during the past several years. This growth is directly related to the growth of the media themselves. Consumers are no longer restricted in their choice of media, but can choose from a large number of magazines, newspapers, radio stations, and (thanks to the cable boom) television channels appealing to a wide variety of interests and tastes. All of these rapid developments have created a need by broadcasters, advertisers, and media critics for information about the media and those who use the media. Research is no longer restricted to head counts of the audience.

This book is designed for the beginning mass media research course. It is divided into five parts. Part One concentrates on the research process and discusses the components of scientific research, the research procedure, and sampling. Part Two presents four approaches to research: the laboratory approach, survey research, field research, and content analysis. Part Three gives an introduction to univariate and multivariate statistical methods. (These chapters are intended as a basic discussion and not as an exhaustive study of all available statistical methods; interested readers should refer to more specialized statistics books.) Part Four discusses three major applications of research: the print media, the electronic media, and advertising and public relations. Part Five provides information on analyzing data and reporting results.

Each chapter concludes with problems and questions for further consideration, and an appendix of tables and a glossary appear at the end of the book (terms included in the glossary are printed in **boldface** when they are introduced in the text). Readers are encouraged to consult the various trade papers and academic journals in mass media to keep up to date with developments.

We wish to acknowledge the special contributions made by two reviewers who read the manuscript at all stages: Alan M. Rubin, Kent State University; and James R. Smith, SUNY–New Paltz. We also thank the following reviewers who offered suggestions at one or more stages: David W. Clark, Christian Broadcast Network; Susan Tyler Eastman, Indiana University; David Eshelman, Central Missouri State University; Allan Mussehl, Middle Tennessee State University; and Joseph C. Philport, The Arbitron Company. Three colleagues from the University of Georgia also provided valuable assistance: James E. Fletcher offered many helpful suggestions as well as the raw data for our research examples; Leonard Reid and Tom Russell made several suggestions about the manuscript. In addition, James B. Weaver III, a graduate student, provided the computer program materials used in Chapter 16 as well as the random numbers table and the random telephone number computer program (TELEGEN). Another student, Lanie Pryles, graciously contributed a great deal of time and effort on the glossary and in reading page proofs.

We also wish to thank our editor, Becky Hayden, for her expertise and advice throughout the project. She is a consummate professional. Our production editor, Sally Schuman, was instrumental in getting this project completed. And our copy editor, William Reynolds, deserves a great deal of thanks for making the manuscript more readable.

Lastly, as is usually the case with co-authored works, if this book contains any errors, inaccuracies, omissions, or other lapses, each author will steadfastly blame the other.

Roger Wimmer
Joseph Dominick

CONTENTS IN BRIEF

CONTENTS

PART 1

The Research Process

CHAPTER 1

Science and Research

Several years ago Richard Weaver (1953), a communication scholar, discussed the meanings of language and identified the differences between "god" and "devil" terms. A god term is positive and has connotations of strength, goodness, and significance; *democracy, innovation,* and *freedom* are god terms in the United States. A devil term, on the other hand, represents a negative image and connotes weakness, evil, or impending doom; *communist, moral decay,* and *inferior* are examples of devil terms.

One term that transcends both categories is *research.* Advertisers, for example, use research as a god term to sell products and services. Broadcast commercials and print advertisements include statements such as: "Research shows that six out of ten doctors . . ." or "According to a recent survey, eighty out of one hundred Cadillac owners preferred. . . ." The intention is to associate with the product a degree of importance on the basis of what the researchers have concluded; research results alone are considered enough to convince consumers of the need for a product.

Research can also be a devil term, however, especially to those mass media students who consider statistics and research to be detours in the road to receiving a college degree. It is the intention of this book to help dispel the "devil" connotation of research. Communication research need not be viewed negatively; rather it should be regarded as a tool with which to search for answers.

Chapter 1 provides an introduction to mass media research and includes discussions of the development of mass media research during the past forty years, the methods used in collecting and analyzing information, and an expanded discussion of the scientific method of research. This chapter provides the foundation for topics discussed in greater detail in later chapters.

Two basic questions that the beginning researcher must learn to answer are *how* and *when* to use statistical procedures and research methods. Developing methods and procedures are valuable tasks, but the focus for the majority of research students should be on applications. This book advocates the approach of the *applied data analyst,* not the statistician.

As Figure 1.1 shows, the statistician and the data analyst belong on the same continuum, but their specialties are different. Statisticians generate statistical procedures or formulas called **algorithms;** data analysts use these algorithms to investigate research questions and hypotheses. The results of this cooperative effort are used to advance our understanding of the mass media.

For example, consider the problem broadcasters once had with statistical error in audience ratings (a problem which has not yet been totally solved). Broadcasters and advertisers who used these ratings recognized that due to sampling procedures, the figures they obtained were highly susceptible to error. The users of ratings approached statisticians and asked them to develop a sampling error formula that would help make ratings more reliable. Several formulas were developed, and with these formulas, researchers, broadcasters, and advertisers were able to estimate the amount of error present in ratings data and help make more reliable interpretations.

Since the early part of the twentieth century, when it required little more than "head counts" of the audience, the mass media industry has come to rely on research results for nearly every major deci-

Figure 1.1 Research continuum

sion it makes. This increased demand for information has created a need for more research organizations, both public and private, as well as an increase in research specialization. There are research directors who plan and supervise studies and act as liaisons to management; methodological specialists who provide statistical support; research analysts who design and interpret studies; and computer specialists who provide hardware and software support in data analysis.

The importance of research in mass media is partly due to the realization that "gut" feelings or reactions are not entirely reliable as bases for decisions. Although common sense is often accurate, media decision makers need additional, more objective information in order to evaluate a problem, especially when hundreds of thousands of dollars are at stake. Thus, the last fifty years have witnessed the slow evolution of an approach combining research and intuition to create a higher probability of success in the decision-making process.

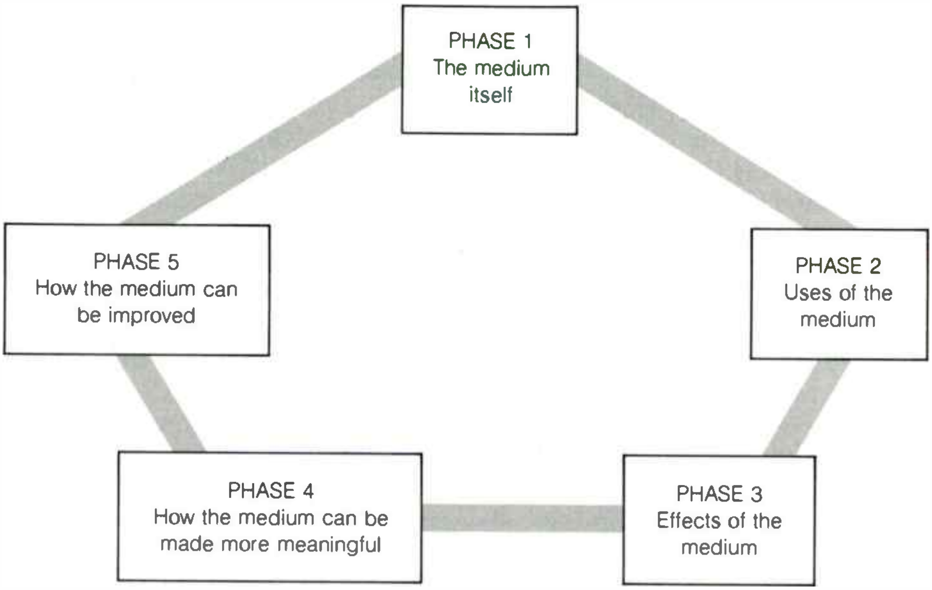
Research is not limited only to decision-making situations. It is also widely used in theoretical areas to attempt to describe the media, analyze media effects on consumers, and so forth. Barely a day goes by without some reference in the media to audience surveys, public opinion polls, the status or growth of one medium or another, or the success of advertising or public relations campaigns. Research in all areas of the media is expanding at a phenomenal rate.

The Development of Mass Media Research

Mass media research has evolved in definable steps, and each medium's requirements for research have followed a similar pattern, as illustrated in Figure 1.2. Initially, there is an interest in the medium itself: researchers attempt to explain what the medium is, how it developed, and how it offers alternatives to the communication systems already available. This initial interest is followed by a second phase: a desire to learn more about how people use the medium for information and entertainment. A third phase involves an interest in determining what effects the medium has on people—how the medium alters or contributes to behavior patterns and attitudes.

Generally speaking, the interest in effects never subsides, but other phases quickly evolve. One is an investigation of how to structure the medium to make it more useful or significant in the world of communication (see *Journal of Broadcasting* 24:3, 1979) and includes the broad area of policy research. Another is an interest in the effects of technological developments on the importance of the medium (Levy 1980; Agostino, Terry, and Johnson 1980). These latter two phases are clearly evident in studies of cable television's impact on the television industry. As Figure 1.2 indicates, mass media research is not a linear process: research emphases are continually changing. Researchers continue to apply all phases of research to a medium even decades after it has become popular.

Figure 1.2 Research phases in mass media



At least four major events or social forces have contributed to the growth of mass media research. The first was World War I, which brought about a need to further understand the nature of propaganda. Researchers working from a stimulus-response point of view attempted to uncover the effects of the media on people (Lasswell 1927). The media at that time were thought to exert a very powerful influence over their audiences, and several assumptions were made about what the media could and could not do. One theory of mass media was known as the hypodermic needle model of communication, which suggested that a mass communicator need only "shoot" messages at an audience and receive preplanned and universal effects. The belief was that all people behave in very similar ways when they encounter media messages, though we know now that individual differences among people rule out this rather simplistic view of the media. However, as DeFleur and Ball-Rokeach (1975) note:

> These assumptions may not have been explicitly formulated at the time, but they were drawn from fairly elaborate theories of human nature, as well as the nature of the social order. It was these theories that guided the thinking of those who saw the media as powerful.

A continuation of this first social force was during the 1930s and 1940s, when media researchers realized that early theories and assumptions about the media were limited in scope and application (see DeFleur and Ball-Rokeach 1975). Instead of continuing to investigate other phases of the media, researchers stepped back to reexamine the media themselves—to explain the basic characteristics of the media. This was soon followed by an interest in the uses and effects of the media (Katz and Lazarsfeld 1955; Lazarsfeld, Berelson, and Gaudet 1948; Herzog 1944; Klapper 1960).

A second contributor to the development of mass media research was the realization by advertisers in the 1950s and 1960s that research data were useful in persuading potential customers to buy products and services. Consequently, they encouraged studies of message effectiveness, audience demographics and size, placement of advertising to achieve the highest level of exposure (efficiency), frequency of advertising necessary to persuade potential customers, and selection of the medium that offered the best chance of reaching the target audience.

A third contributing social force was the increasing interest of citizens in the effects of the media on the public, especially on children. The direct result was an interest in research related to violence and sexual content in television programs and in commercials aired during children's programs. Over the past several years, some researchers have expanded their focus to include the positive (prosocial) as well as the negative (antisocial) effects of television (see *Journal of Broadcasting* 25:4, 1981).

Increased competition among the media for advertising dollars has been a fourth contributor to the growth of research. Media management has grown more sophisticated of late, utilizing long-range plans, management by objectives, and increasing dependency on data to support the decision process. Even program producers have begun to seek relevant research data, a province usually assigned to the creative side of program development. In addition, the mass media are now headed full speed into audience fragmentation; we are now experiencing a "de-massification" of the mass media.

The competition among the media for audiences and advertising dollars continues to reach new levels of complexity. The media "survival kit" today includes information about consumers' changing values and tastes, shifts in demographic patterns, and developing trends in lifestyles. Audience fragmentation in the media has led to an increased desire for trend studies (fads, new behavior patterns), image studies (how people perceive the media and their environment), and segmentation studies (explanations of types or groups of people). Major research organizations, consultants, and media owners and operators now conduct research that was previously considered the sole property of marketing. With the advent of increased competition and audience fragmentation, the media are more frequently using marketing strategies in an attempt to discover their position in the marketplace, and when this position is isolated or identified, the medium is packaged as an "image" rather than a product. (Similarly, the producers of consumer goods such as soap and toothpaste try to sell the "image" of these products, since the products themselves are very similar, if not the same, from company to company.) This packaging strategy involves determining what the members of the audience think, how they use language, how they occupy their spare time, and so on. Information on these ideas and behavior patterns is then woven into the merchandising effort to make the medium seem to be part of the audience. Positioning thus involves taking information from the audience, transforming it into research data, and using it to market the medium (see Ries and Trout 1981 for more information about positioning in broadcasting and advertising).

Much of the media research up to the early 1960s originated in psychology and sociology departments at colleges and universities. Researchers with backgrounds in the media were rare because the media themselves were young. But this situation has changed. Media departments in colleges and universities grew rapidly in the 1960s, and media researchers entered the scene. Today the field is no longer dominated by researchers from allied areas. In fact, the trend is to encourage cross-disciplinary studies in which media researchers invite participation from sociologists, psychologists, and political scientists. Because of the pervasiveness of the media, researchers from all areas of science are now actively involved in attempting to answer media-related questions.

Media Research and the Scientific Method

Kerlinger (1973) defined scientific research as a systematic, controlled, empirical, and critical investigation of hypothetical propositions about the presumed relations among natural phenomena. Regardless of its origin, all research begins with a basic question or proposition about some phenomenon; for example: Why do people watch television? What section of the newspaper do people read most often? Why don't people subscribe to the most expensive cable television tier? Each of these questions could be answered to some degree with a research study.

There are several possible approaches in answering research questions. Kerlinger (1973) labeled the methods of finding answers "methods of knowing" and identified them as tenacity, intuition, authority, and science.

The **method of tenacity** follows the logic that "something is true because it has always been true." An example is the store owner who says, "I don't advertise because my parents did not believe in advertising." The basic idea is that nothing changes; what was good, bad, or successful before will continue to be so in the future.

The **method of intuition,** or a priori approach, assumes that something is true because it is "self-evident" or "stands to reason." A person who claims that "no one likes *Playboy* magazine because it contains pictures of nude women" is using the method of intuition.

The person assumes that because he or she does not like the magazine (or television program, or newspaper editorial), everyone else must hold the same belief.

The **method of authority** involves believing something because a trusted source, such as a relative, news correspondent, or teacher, says it is true. The emphasis is on the source and not on the methods the source may have used to gain the information. A person who claims that "the world is going to end tomorrow because the *New York Times* editorial said so" is operating according to the method of authority.

The **scientific method** approaches learning as a series of small steps. That is, one study or one source provides only an *indication* of what may or may not be true; the "truth" is found only through a series of objective analyses. This means that the scientific method is self-correcting in that changes in thought or theory are appropriate when errors in previous research are uncovered. For example, scientists changed their ideas about the planet Saturn when, on the basis of information gathered by the Voyager space crafts, they uncovered errors in earlier observations. In communication, researchers discovered that the early perceptions of the power of the media (the "hypodermic needle" theory) were incorrect and, after numerous research studies, concluded that behavior and ideas are changed by a combination of communication sources and that people may react to the same message in different ways.

The scientific method may be inappropriate in many areas of life, such as evaluating works of art, choosing a religion, or forming friendships, but the method has been valuable in producing accurate and useful data in mass media research. The following section provides a more detailed look at this method of knowing.

Characteristics of the Scientific Method

The scientific method embodies five basic characteristics or tenets that distinguish it from other methods of knowing. Any research approach not following these tenets cannot be considered a scientific approach.

Scientific research is public. Scientific advancement depends on freely available information. A researcher (especially in the academic sector) cannot plead private knowledge, methods, or data in arguing for the accuracy of his or her findings; scientific research information must be freely communicated from one researcher to another. As Nunnally (1978:8) noted:

> Science is a highly public enterprise in which efficient communication among scientists is essential. Each scientist builds on what has been learned in the past, day by day his or her findings must be compared with those of other scientists working on the same types of problems.

Researchers, therefore, must take great care in published reports to include information on their use of sampling methods, measurements, and data gathering procedures. This information allows other researchers to independently verify, or replicate a given study and to support or refute the initial research findings. **Replication,** discussed in greater detail in Chapter 3, allows for correction or verification of previous research findings.

Researchers also need to save their descriptions of observations (data) and research materials so that information not included in a formal report can be made available to other researchers upon request. It is common practice to keep all raw research material for a period of five years. This material is usually provided free as a courtesy to other researchers, or for a nominal fee if photocopying or additional materials are required.

Science is objective. Science tries to rule out eccentricities of judgment by researchers. When a study is undertaken, explicit rules and procedures are constructed and the researcher is bound to follow them; let the chips fall where they may. Rules for classifying behavior are used so that two or more independent observers can classify particular patterns of behavior in the same manner. For example, if the attractiveness of a television commercial is being measured, researchers might count the number of times a viewer switches channels while the commercial is shown. This is considered an objective measure because any observer would report a change in channel. Conversely, to measure attractiveness by observing how many people make "negative facial expressions" while the ad is shown would be subjective, since observers might have different ideas of what constitutes a negative expression. However, an explicit definition of the term "negative facial expression" might eliminate the coding error.

Objectivity also requires that scientific research deal with facts rather than interpretations of facts. Science rejects its own authorities if their statements are in conflict with direct observation. As the noted psychologist B. F. Skinner (1953) wrote: "Research projects do not always come out as one expects, but the facts must stand and the expectations fall. The subject matter, not the scientist, knows best."

Science is empirical. Researchers are concerned with a world that is knowable and potentially measurable ("empiricism" is derived from the Greek word for "experience"). They must be able to perceive and classify what they study and to reject metaphysical and nonsensical explanations of events. For example, a newspaper publisher's claim that declining subscription rates are "God's will" would be rejected by scientists—such a statement cannot be perceived, classified, or measured.

This does not mean that scientists evade abstract ideas and notions—they encounter them every day. But they recognize that concepts must be strictly defined to allow for observation and measurement. Scientists must designate certain kinds of overt behavior patterns to represent abstract concepts in order to provide a link between the metaphysical and the physical world. Typically this linkage is accomplished by framing an **operational definition.**

An operational definition is important in science and necessitates a brief introduction. There are basically two kinds of definitions. A **constitutive definition** defines a word by substituting other words or concepts for it. For example, "An artichoke is a green leafy vegetable, a tall composite herb of the *Cynara scolymus* family" is a constitutive definition of the concept "artichoke." In contrast, an operational definition specifies procedures to be followed in order to experience or measure a concept. For example, "Go to the grocery store and find the produce aisle. Look for a sign that says 'Artichokes.' What's underneath the sign is one." Although an operational definition assures precision, it does not guarantee validity. An errant stock clerk may mistakenly stack lettuce under the artichoke sign and fool someone. This underlines the importance of considering both the constitutive and the operational definition of a concept in evaluating the trustworthiness of any measurement. A careful examination of the constitutive definition of artichoke would indicate that the operational definition might be faulty. (See Chapter 2 and Nunnally 1978 for further discussion of operational definitions.)

Science is systematic and cumulative. No single research study stands alone, nor does it rise or fall by itself. Astute researchers always utilize previous studies as building blocks for their own work. One of the first steps taken in conducting research is to review the available scientific literature on the topic so that the current study will draw upon the heritage of past research (see Chapter 3). This review is valuable for identifying problem areas and important factors that might be relevant to the current study (see Cattell 1966).

In addition, scientists attempt to search for order and consistency among their findings. In its ideal form, scientific research begins with a single, carefully observed event and progresses ultimately to the formulation of theories and laws. A **theory** is a set of related propositions that presents a systematic view of phenomena by specifying relationships among concepts. Researchers develop theories by searching for patterns of uniformity to explain the data that have been collected. In situations where relationships among variables are invariant under given conditions (the relationship is always the same), researchers may formulate a law. Both theories and laws help researchers search for and explain consistency in behavior, situations, and phenomena.

Science is predictive. Science is concerned with relating the present to the future. In fact, one reason why scientists strive to develop theories is that they are useful in predicting behavior. A theory's adequacy lies in its ability to predict a phenomenon or event successfully. If a theory suggests predictions that are not borne out by an analysis of available data, then that theory must be carefully reexamined and perhaps discarded. Conversely, if a theory generates predictions that are supported by the data, then that theory can be used to make predictions in other situations.

Research Procedures

The use of the scientific method of research is intended to provide an objective, unbiased evaluation of data. In order to systematically investigate research questions and hypotheses, both academic and private sector researchers follow a basic seven-step developmental chain of procedures:

1. Problem development
2. Review of previous literature in the area
3. Research design
4. Data collection
5. Data analysis
6. Interpretation and conclusions
7. Replication

Each step is dependent upon all others to help produce a maximally efficient research study. Before a literature search is possible, a clearly stated research problem is required; in order to design the most efficient method of investigating a problem, the researcher needs to know what types of studies have been conducted, and so on. All of the steps are interactive: the results or conclusions of any step have a bearing on other steps. For example, the literature search may refine and sometimes alter the initial research problem.

Two Sectors of Research: Academic and Private

In discussions of research, most people identify two major sectors, academic and private. Both are important in mass media research, and in many cases both work together to solve media problems.

Academic sector research usually implies research that is conducted by academicians. It also generally means that the research has a *theoretical* approach; that is, the results are intended to help explain the mass media and their effects on individuals. Some popular research topics in the theoretical area include the use of the media and various media-related items, such as video games, teletext, and satellite distribution of programs; lifestyle analyses of consumers; media "overload" on consumers; alternatives to present media systems; and the effects of violent content and commercials on children.

Private sector research usually implies research that is conducted by private industries or their research consultants. It also means that the research generally has an *applied* nature; that is, the results are intended to be used in decision-making situations. Typical research topics in the private sector include analyses of media content and consumer preferences, advertising (frequency, length, scheduling, recall), marketing (media mixing), public relations approaches and techniques, and sales forecasting.

There are other differences between academic and private sector research. For instance, academic research is public. Any other researcher or research organization that wishes to use the information gathered by academic researchers can do so (ideally). Private sector research, on the other hand, is known as **proprietary data:** it is considered the sole property of the sponsoring agency and cannot generally be obtained by other researchers. Some private sector research is released after several years, but this practice is the exception rather than the rule.

Also, academic research is generally less expensive to conduct than research in the private sector. This is not to say that academic research is "cheap"—it is not in many cases. But academicians do not need to have enormous sums of money to cover overhead costs for equipment, facilities, subcontractors, and personnel. Private sector research, whether it is done within a company or hired out to a research firm, must take such expenses into account. This is the primary reason why the major networks and other large media organizations sometimes prefer to use academic researchers rather than professional research firms.

Despite their differences, it is important for beginning researchers to understand that academic research and private sector research are not completely independent of each other. The link between the two areas is important. Academicians perform many studies for industry, and private sector groups conduct research that can be classified as theoretical (for example, the American Broadcasting Company has a Department of Social Research). Many college and university professors act as consultants to, and often conduct private sector research for, the media industry.

It is also important for all researchers to refrain from attaching stereotypical labels to academic or private sector research: unrealistic, inappropriate, pedantic, limited in scope, and so on. Research in both sectors, although differing occasionally in terms of cost and scope, uses similar methodologies and statistical analyses. In addition, both sectors have common research goals: to understand and to predict.

In conducting a study according to the scientific method, researchers need to have a clear understanding of what they are investigating, how the phenomenon can be measured or observed, and what procedures are required to test the observations or measurements. Conceptualization of the research problem in question and a logical development of procedural steps are necessary in order to have any hope of answering a research question or hypothesis. (See Chapter 3 for a more detailed discussion of research procedures.)

Summary

In an effort to understand any phenomenon, researchers can follow one of several methods of inquiry. Of the procedures discussed in this chapter, the scientific approach is most applicable to the mass media because it involves a systematic, objective evaluation of information. Researchers first identify a problem, then investigate it, using a prescribed set of procedures known as the scientific method of research. In addition, the scientific method is the only learning approach that allows for self-correction of research findings; one study does not stand alone but must be supported or refuted by others.

The rapid growth of mass media research during the past several years is mainly attributable to the rapidly developing technology of the media industry. Because of this growth in research, both applied and theoretical approaches have taken on more significance in the decision-making process of the mass media and in our understanding of the media.

Questions and Problems for Further Investigation

1. Obtain a recent issue of *Journalism Quarterly, Journal of Broadcasting,* or *Public Opinion Quarterly.* How many articles fit into the research phases outlined in Figure 1.2?
2. What are some potential research questions that might be of interest to both academic and private sector researchers? Do not limit the questions to the area of effects research.
3. How might the scientific research approach be abused by researchers?
4. Theories are important in developing solid bodies of information; they are used as springboards to investigation. However, there are few universally recognized theories in mass media research. Why do you think this is true?
5. During the past several years, citizens' groups have claimed that television has a significant effect on viewers, especially as regards violence and sexual content of programs. How might these groups have collected data to support these claims? Which method of knowing would such citizens' groups be most likely to use?

References and Suggested Readings

Agostino, D. E., Terry, H. A., and Johnson, R. C. 1980. "Home Video Recorders: Rights and Ratings." *Journal of Communication* 30:4, 28–35.

Brown, J. A. 1980. "Selling Airtime for Controversy: NAB Self Regulation and Father Coughlin." *Journal of Broadcasting* 24:2, 199–224.

Carroll, R. L. 1980. "The 1948 Truman Campaign: The Threshold of the Modern Era." *Journal of Broadcasting* 24:2, 173–88.

Cattell, R. B., ed. 1966. *Handbook of Multivariate Experimental Psychology.* Chicago: Rand McNally.

DeFleur, M. L. and Ball-Rokeach, S. 1975. *Theories of Mass Communication.* New York: David McKay.

Herzog, H. 1944. "What do We Really Know about Daytime Serial Listeners?" *Radio Research 1943–44,* edited by P. Lazarsfeld and F. Stanton. New York: Duell, Sloan, and Pearce.

Katz, E. and Lazarsfeld, P. F. 1955. *Personal Influence.* Glencoe: Free Press.

Kerlinger, F. N. 1973. *Foundations of Behavioral Research.* New York: Holt, Rinehart and Winston.

Klapper, J. 1960. *The Effects of Mass Communication.* New York: Free Press.

Lasswell, H. D. 1927. *Propaganda Technique in the World War.* New York: Alfred A. Knopf.

Lazarsfeld, P., Berelson, B., and Gaudet, H. 1948. *The People's Choice.* New York: Columbia University Press.

Levy, M. R. 1980. "Home Video Recorders: A User Survey." *Journal of Communication* 30:4, 23–27.

Murphy, J. H. and Amundsen, M. S. 1981. "The Communication Effectiveness of Comparative Advertising for a New Brand on Users of the Dominant Brand." *Journal of Advertising* 10:1, 14–20.

Nunnally, J. C. 1978. *Psychometric Theory.* New York: McGraw-Hill Book Company.

Ries, E. and Trout, J. 1981. *Positioning: The Battle for Your Mind.* New York: McGraw-Hill Book Company.

Skinner, B. F. 1953. *Science and Human Behavior.* New York: Macmillan.

Sybert, P. J. 1980. "MBS and the Dominican Republic." *Journal of Broadcasting* 24:2, 189–98.

Weaver, R. M. 1953. *The Ethics of Rhetoric.* Chicago: Henry Regnery.

CHAPTER 2

Elements of Research

Chapter 1 presented a brief overview of the research process. In this chapter, three basic elements of this process are defined and discussed: (1) concepts and constructs, (2) variables, and (3) measurement. It is necessary to understand these elements in order to conduct and understand empirical research.

Concepts and Constructs

A **concept** is a term that expresses an abstract idea formed by generalization from particulars (Kerlinger 1973). It is formed by summarizing related observations. For example, a researcher might observe that a public speaker becomes restless, starts to perspire, and continually fidgets with a pencil just before giving an address. The researcher might summarize these observed patterns of behavior and label them *speech anxiety*. On a more concrete level, the word *table* is a concept that summarizes a wide variety of observable objects, ranging from a plank supported by concrete blocks to a piece of furniture typically found in dining rooms. In mass communication, terms such as *mes-*

sage length, *media usage*, and *readability* might typically be used as concepts.

Concepts are important because they facilitate communication among those who have a shared understanding of them. Researchers use concepts to organize their observations into meaningful summaries and to transmit this information to their colleagues and to the public. In addition, since concepts are abstracted from observations, they enable researchers to look for general explanations or patterns in these observations.

A **construct** is a special type of concept created for a particular scientific purpose and consists of a combination of other concepts. Constructs are generally difficult to observe directly; their existence must be inferred from related behavior patterns. For example, in mass communication research, one occasionally encounters the term *authoritarianism*. This is a construct specifically defined to describe a certain type of personality; it comprises nine different concepts, including conventionalism, submission, superstition, and cynicism. Authoritarianism itself cannot be seen; its presence must be determined by some type of questionnaire or standardized test. The results of such tests provide indications of what authoritarianism might be and if it exists, but they do not provide exact definitions for the concept.

A **variable** is a phenomenon or event that can be measured or manipulated and is used in the development of constructs. Researchers attempt to test a number of associated variables in order to develop an underlying meaning or relationship among them. After suitable analysis, the important variables are retained while the others are discarded. These important variables are labeled **marker variables** since they seem to define or highlight the construct under study. After further analysis, new marker variables may be added to increase understanding of the construct and to permit more reliable predictions.

Constructs and marker variables are valuable tools in theoretical research. But, as noted in Chapter 1, researchers also function at the observational or empirical level. In order to understand how this is done, it is necessary to examine variables and how they are measured.

Independent and Dependent Variables

Variables are classified in terms of their relationship with one another. It is customary to talk about *independent* and *dependent* variables: independent variables are systematically varied by the researcher, while dependent variables are observed and their values presumed to depend on the effects of the independent variables. In other words, the dependent variable is what the researcher wishes to explain. For example, assume an investigator is interested in determining how the angle of a camera shot affects an audience's perception of the credibility of a television newscaster. Three different versions of a newscast are videotaped: one shot from a very low angle, another from a high angle, and a third from eye level. Groups of subjects are randomly assigned to view one of the three versions and to complete a questionnaire that measures credibility. In this experiment, the camera angle is the independent variable. Its values are systematically varied by the experimenter who selects only three of the possible camera angles available. The dependent variable to be measured is the perceived credibility of the newscaster. If the researcher's assumption is correct, then the newscaster's credibility will vary according to the camera angle. (Note that the actual values of the dependent variable are not manipulated; they are simply observed.)

Keep in mind that the distinction between types of variables depends on the purposes of the research. An independent variable in one study may be a dependent variable in another. Also, there are times when a research task involves examining the relationship of more than one independent variable to a single dependent variable. For example, a study designed to examine the impact of type size and page layout on learning would encompass two independent variables —type size and layout—and one dependent variable—learning. Moreover, there are many instances where multiple dependent variables are measured in a single study. This type of study is called a *multivariate analysis* and will be discussed in Chapter 12.

Other Types of Variables

In nonexperimental research, where there is no active manipulation of variables, different terms are sometimes substituted for independent and dependent variables. The variable that is used for predictions or assumed to be causal (analogous to the independent variable) is sometimes called the **predictor** or **antecedent variable.** The variable that is predicted or assumed to be affected (analogous to the dependent variable) is sometimes called the **criterion variable.**

Researchers often wish to account for or control certain types of variables for the purpose of eliminating unwanted influences. These **control variables** are used to ensure that the results of the study are due to the independent variables and not some other source. However, a control variable need not always be used to eliminate an unwanted influence. On occasion, researchers use a control variable such as age, sex, or socioeconomic status to divide subjects into specific relevant categories. For example, in studying the relationship between newspaper readership and reading ability, it is apparent that IQ will affect the relationship and must be controlled; thus, subjects may be selected on the basis of IQ scores, or placed in groups with similar scores. One problem that researchers often encounter is specifying all the variables that must be controlled in order to avoid spurious results.

Operationally Defining Variables

In Chapter 1 it was stated that an operational definition specifies procedures to be followed in order to experience or measure a concept. Research depends upon observations, and observations cannot be made without a clear statement of what is to be observed. An operational definition is such a statement.

Operational definitions are indispensable in scientific research because they enable investigators to measure relevant variables. In any study, it is necessary to provide operational definitions for both independent and dependent variables. Figure 2.1 contains examples of such definitions taken from research studies in mass communication.

Kerlinger (1973) identified two types of operational definitions, measured and experimental. A measured operational definition specifies how to measure a variable. For instance, a researcher investigating dogmatism and media use might operationally define *dogmatism* as a subject's score on the Twenty-Item Short Form Dogmatism Scale. An experimental operational definition explains how an investigator has manipulated a variable. Obviously, this type of definition is used when defining the independent variable in a laboratory setting. For example, in a study concerning the impact of television violence, the researcher might manipulate media violence by constructing two eight-minute films. The first film, labeled "the violent condition,"

Figure 2.1 Illustrations of operational definitions

Study	*Variable*	*Operational Definition*
Elliot and Slater (1980)	Perceived TV Program Realism	Subjects evaluated a program on a five-point scale ranging from "very realistic" to "very unrealistic."
Levy (1979)	Gregariousness	Ranking was based on each subject's total number of friendships and memberships in voluntary organizations.
Rossiter and Robertson (1980)	Exposure to TV Drug Advertising	The number of proprietary drug commercials per program, as obtained from Broadcast Advertising Reports, was weighted according to viewing frequencies as reported by subjects.
Baran, Chase, and Courtright (1979)	Prosocial Behavior	Subjects were rated according to whether they retrieved an armload of books dropped by the experimenter's confederate and how much time elapsed before they responded.
Burgoon and Burgoon (1980)	Time Spent Reading Newspaper	Subjects estimated the amount of time they spent yesterday reading the paper.

might contain scenes from a boxing match. The second film, labeled "the non-violent condition," could depict a swimming race. Or, to take another example, source credibility might be manipulated by alternately attributing an article on health to the *New England Journal of Medicine* and to the *National Enquirer* in order to determine which source has the higher credibility.

Operationally defining a variable forces the researcher to express abstract concepts in concrete terms. Occasionally, after unsuccessfully grappling with the task of making a key variable operational, the investigator may conclude that the variable as originally conceived is too vague or ambiguous and that redefinition is required. Because operational definitions are expressed so concretely, they can communicate *exactly* what the terms represent. For instance, a researcher might define political knowledge as the number of correct answers on a twenty-item true/false test. While it is possible to argue over the validity of this particular operational definition, there is no confusion as to what is meant by the statement, "Women possess more political knowledge than men."

Finally, there is no single infallible method for operationally defining a variable. No operational definition satisfies everybody's idea of what it was intended to represent. An investigator must decide which method is best suited for the research problem at hand.

Measurement

Mass communication research, like all research, can be qualitative or quantitative. **Qualitative research** describes or analyzes a phenomenon without specifically measuring variables. No statistical analysis is involved in qualitative research, although the data might be expressed numerically. Some common forms of qualitative research are legal research, historical research, and critical research. Qualitative research is also used frequently in the mass media. Newcombe's book *Television: The Critical View* (1979) is a collection of qualitative articles. Other qualitative research studies are Thorburn's (1981) analysis of television melodrama and Davidson's (1981) treatment of television docudramas.

Quantitative research requires that the variables under consideration be measured. This form of research is concerned with how often a variable is present and generally uses numbers to communicate this amount. Quantitative research has certain advantages. One is that the use of numbers allows greater precision in reporting results. For example, the Violence Index (Gerbner, *et al.* 1980), a quantitative measuring device, makes it possible to report the exact increase or decrease in violence from one television season to another, whereas qualitative research could only describe whether violence went up or down. Another advantage is that quantitative research permits the use of powerful methods of mathematical analysis. The importance of mathematics to mass media research is difficult to overemphasize. As pointed out by measurement expert J. P. Guilford (1954):

> The progress and maturity of a science are often judged by the extent to which it has succeeded in the use of mathematics. . . . Mathematics is a universal language that any science or technology may use with great power and convenience. Its vocabulary of terms is unlimited . . . . Its rules of operation . . . are unexcelled for logical precision.

This book is concerned with the procedures and skills relevant to quantitative research. This is not to say that quantitative research is in some sense "better" than qualitative research; each technique has value. Certain situations lend themselves more readily to one technique than the other. Over the past thirty years, however, quantitative research has become more and more common in mass media. Consequently, it is increasingly important for beginning researchers to familiarize themselves with common quantitative techniques.

The Nature of Measurement

The idea behind **measurement** is a simple one: it is a procedure in which a researcher assigns numerals to objects, events, or properties according to certain rules. Examples of measurement are everywhere: "She or he is a 10." "Unemployment increased by 1%." "The earthquake measured 5.5 on the Richter scale." Note that the definition contains three central concepts: numerals, assignment, and rules. A numeral is a symbol, such as V, X, C, or 5, 10, 100. A numeral has no implicit quantitative meaning. When it is given quantitative meaning, it becomes a number and can be used in mathematical and statistical computations. Assignment is the designation of numerals or numbers to certain objects or events: A simple measurement system might entail assigning the numeral "1" to those people who get most of their news from television, the numeral "2" to those get most of their news from a newspaper, and the numeral "3" to those who get most of their news from some other source.

Rules specify the way that numerals or numbers are to be assigned. Rules are at the heart of any measurement system; if they are faulty, so will be the system. In some situations, the rules are obvious and straightforward. To measure reading speed, a stopwatch and a standardized message may be sufficient. In other instances, the rules are not so apparent. Measuring certain psychological traits such as "source credibility" or "attitude toward violence" calls for carefully explicated measurement techniques.

Additionally, in mass communication research and in much of social science research, investigators usually measure indicators of the properties of individuals or objects rather than the individuals or objects themselves. Concepts such as "authoritarianism" or "motivation for reading the newspaper" cannot be directly observed; they must be inferred from presumed indicators. Thus, if a person endorses statements such as "Orders from a superior should always be followed without question" and "Law and order are the most important things in society," it can be deduced that he or she is more authoritarian than someone who disagreed with those statements.

Lastly, measurement systems strive to be isomorphic to reality. **Isomorphism** means identity or similarity of form or structure. In order for the mass communication researcher to perform certain operations with numbers, the structure of the measurement system used to assign them to observations should be isomorphic to some numerical structure. In some research areas, such as the physical sciences, isomorphism is not a key problem, since there is usually a direct relationship between the objects being measured and the numbers assigned to them. For example, if an electrical current travels through substance A with less resistance than it does through substance B, then it can be deduced that A is a better conductor than B. Testing a few more substances can lead to a ranking of conductors whereby the numbers assigned indicate the degree of conductivity. The measurement system is isomorphic to reality.

In mass media research, the correspondence is seldom that obvious. For example, imagine that a researcher is trying to develop a scale to measure the "persuasibility" of people in connection with a certain type of advertisement. She devises a test and administers it to five people. The scores are displayed in Figure 2.2. Now imagine that an omniscient being is able to disclose the "true" persuasability of those same five people. These scores are also shown in Figure 2.2. For two people, the test scores correspond exactly to the "true" scores. The other three scores miss the "true" scores, but there is a correspondence between the rank orders. Also note that the "true" persuasability scores ranged from 0 to 12, while the measurement scale ranged from 1 to 8. To summarize, there is a general correspondence between the test and reality, but the test is far from an exact measure of what actually exists.

Unfortunately, the degree of correspondence between measure-

Figure 2.2 Illustration of isomorphism

Person	Test Score	Real Score
A	1	0
B	3	1
C	6	6
D	7	7
E	8	12

ment and reality is rarely known in actual research. In some cases researchers are not even sure they are actually measuring what they are trying to measure. In any event, researchers must carefully consider the degree of isomorphism between measurement and reality. This topic is discussed in greater detail later in the chapter.

Levels of Measurement

Scientists have distinguished four different ways to measure things, or four different levels of measurement. The operations that can be performed with a given set of scores depend upon the level of measurement achieved. The four levels of measurement are: (1) nominal, (2) ordinal, (3) interval, and (4) ratio.

The **nominal level** is the weakest form of measurement. In nominal measurement, numerals or other symbols are used to classify persons, objects, or characteristics. For example, in the physical sciences, rocks can generally be classified into three categories: igneous, sedimentary, and metamorphic. A geologist who assigns a "1" to igneous, a "2" to sedimentary, and a "3" to metamorphic forms a nominal scale. Note that numerals are simply labels that stand for a category; they have no mathematical significance. Just because a rock is placed in category "3" does not mean that it has more "rockness" than those in categories "2" and "1." Other examples of nominal measurement are numbers on football jerseys, social security numbers, and license plates.

The nominal level, like all levels, possesses certain formal properties. Its basic property is that of equivalence. If an object is placed in category "1," it is considered equal to all other objects in that category. To illustrate, suppose a researcher is attempting to classify all the advertisements in a magazine according to their primary appeal. If the ad has an economic appeal it is placed in category "1"; if it uses an appeal to fear, it is placed in category "2"; and so on. Note that all ads using "fear appeal" are considered equal even though they may differ on other dimensions, such as product type, size, or use of illustrations.

Another property of nominal measurement is that all categories are exhaustive and mutually exclusive. This means that each measure accounts for every option possible and that each measurement is appropriate to only one category. For instance, in the previous example of primary appeals in magazine advertisements, all possible appeals would need to be included in the analysis (exhaustive) such as economic, fear, morality, religion, and so on. Each advertisement would be placed in one and only one category (mutually exclusive).

Nominal measurement is frequently used in mass media research. For example, Gafke and Leuthold (1979), in their study of editorial endorsements in the 1976 presidential campaign, divided newspapers into three nominal categories: those that endorsed Ford, those that endorsed Carter, and those that made no endorsement. Wilson and Howard (1979) divided errors found in media reports of news events into two categories: factual errors and interpretative errors.

Even when it is measured at the nominal level, a variable may be used in higher order statistics by *transforming* it into another form. The results of this transformation process are known as **dummy variables.** For example, political party affiliation may be coded as:

Republican	1
Democrat	2
Independent	3
Other	4

This measurement scheme can incorrectly imply that a person classified as "Other" is three units "better" than a person classified as a "Republican." In order to measure political party affiliation and use the data in higher order statistics, the variable must be transformed into a more "equal" form.

One way of transforming the variable to give each option equal value is to recode it as a dummy variable that creates an "either-or" situation for each option: a person is either a "Republican" or something else. For example, a binary coding scheme could be used:

Republican	001
Democrat	010
Independent	100
Other	000

This scheme treats each affiliation equally and allows the variable to be used in higher order statistical procedures.

Note that the final category "other" is coded using all zeros. A complete explanation for this practice is beyond the scope of this book; basically, however, its purpose is to avoid redundancy, since the number of individuals classified as "other" is known from the data on the other options. If, in a sample of a hundred subjects, 25 are found to belong in each of the first three options, it is obvious that there will be twenty-five in the "other" option. (For more information on the topic of dummy variable coding, see Kerlinger and Pedhazur 1978.)

Objects measured at the **ordinal level** are generally ranked along some dimension, usually in a meaningful way, from smaller to greater. For example, one might measure the variable "socioeconomic status" by categorizing families according to class: lower, lower middle, middle, upper middle, or upper. A rank of "1" is assigned to lower, "2" to lower middle, "3" to middle, and so forth. In this situation, the numbers have some mathematical meaning: families in category "3" have a higher socioeconomic status than families ranked "2." Note that nothing is specified with regard to the distance between any two rankings. Ordinal measurement has often been compared to a horse race without a stop watch. The order in which the horses finish is relatively easy to determine, but it is difficult to calculate the difference in time between the winner and the runner-up.

An ordinal scale possesses the property of equivalence (in the previous example, all families placed in any one category are treated equally, even though some might have greater incomes than others). It also possesses the property of *order* among the categories. Any given category can be defined as being higher or lower than any other category. Common examples of ordinal scales are rankings of football or basketball teams, military ranks, restaurant ratings, beauty pageants, and so forth.

Ordinal scales are frequently used in mass communication research. Haskins and Kubas (1979) rated readers' interest in 24 comic strips to see if readership of these comics could be predicted. Williams and Semlak (1979) ranked topics that were important to people during an election campaign to see if they corresponded to the relative degree of media coverage of those issues.

When a scale has all the properties of an ordinal scale and the intervals between adjacent points on the scale are of equal value, then the scale is at the **interval level.** The most obvious example of an interval scale is temperature. The same amount of heat is required to warm an object from 30 to 40 degrees as to warm it from 50 to 60 degrees. Interval scales incorporate the formal property of *equal differences;* that is, numbers are assigned to the positions of objects on an interval scale in such a way that one may carry out arithmetic operations on the differences between thcm.

One disadvantage of an interval scale is that it lacks a true zero point, or a condition of nothingness. For example, it is difficult to conceive of a person having zero intelligence or zero personality. The lack of a true zero point means that the researcher cannot make statements of a proportional nature: someone with an IQ of 100 is not twice as smart as someone with an IQ of 50, and a person who scores 30 on a test of aggression is not three times as aggressive as a person who scores 10.

Interval scales are frequently used in mass communication research. For example, in their study of the impact of the film *All the President's Men,* Elliot and Schenck-Hamlin (1979) constructed a scale to measure attitudes toward the press by measuring responses of "agree" or "disagree" to four separate questions.

Scales at the **ratio level** of measurement have all of the properties of interval scales plus one more: the existence of *true zero point.* With the introduction of this fixed zero point, ratio judgments can be made. For example, since time and distance are ratio measures, one can say that a car traveling at 50 miles per hour is going twice as fast as a car traveling at 25, or that a city lying a hundred miles away is twice as distant as a city 50 miles away. Ratio scales are relatively rare in mass media research, although some variables, such as time spent watching television or number of words per story, are ratio measurements. For example, Gantz (1978) measured news recall ability by asking subjects to report whether they had seen or heard ten items taken from the evening news. Scores could range from zero to 10 on this test.

Researchers using interval or ratio data are able to use statistics specifically designed for these data, procedures known as **parametric statistics.** It should be pointed out that procedures designed for use with "lower" types of data can also be used with data at a higher level of measurement. Statistical procedures designed for higher level data, however, are generally more powerful than those designed for use with nominal or ordinal levels of measurement. Thus, if an investigator has achieved the interval level of measurement, parametric statistics should generally be employed.

There has been some controversy among statisticians as to the importance of the distinction between ordinal and interval scales and the legitimacy of using interval statistics with data that may in fact be ordinal in nature. Without delving too deeply into the arguments involved in this debate, it appears that the safest procedure is to assume interval measurement unless there is clear evidence to the contrary, in which case ordinal statistics should be employed. For example, if a research task involves asking a group of subjects to rank a set of objects, ordinal statistics should be used. If, on the other hand, subjects are given an attitude score constructed by rating responses to various questions, the researcher would be justified in using parametric procedures.

Most statisticians seem to feel that statistical analysis is performed on the numbers yielded by the measures, not the measures themselves, and that the properties of interval scales actually belong to the number system (Roscoe 1975; Nunnally 1978). Additionally, there have been several studies in which various types of data have been subjected to different statistical analyses. These studies suggest that the distinction between ordinal and interval data is not particularly crucial in selecting an analysis method (McNemar 1962).

Discrete and Continuous Variables

There are two forms of variables used in mass media investigations that allow researchers to measure phenomena. A **discrete variable** includes only a finite set of values; it cannot be divided into subparts. For instance, the number of children in a family is a discrete variable because the unit is a person. It would not make much sense to talk about a family size of 2.24 because it is hard to conceptualize .24 of a person. Political affiliation, population, and sex are other discrete variables.

A **continuous variable** can take on any value (including fractions) and can be meaningfully broken into smaller subsections. Height is a continuous variable. If the measurement tool is sophisticated enough, it is possible to have one person 72.113 inches tall and another 72.114 inches tall. Time spent watching television is another example. It is perfectly meaningful to say that person A spent 3.12115 hours viewing while person B watched 3.12114 hours. The *average* number of children in a family is a continuous variable. In this regard, it may be perfectly legitimate to talk about .24 of a person.

When dealing with continuous variables, it is sometimes necessary to keep in mind the distinction between the variable and the measure of the variable. If a child's attitude toward television violence is measured by counting his or her positive responses to six questions, there are only seven possible scores: 0, 1, 2, 3, 4, 5, and 6. It is entirely likely, however, that the underlying variable is continuous even though the measure is discrete. In fact, even if a fractionalized scale were developed, it would still be limited to a finite number of scores. As a generalization, most of the measures in mass communication research tend to be discrete approximations of continuous variables.

Scales and Indexes

Scales and indexes represent composite measures of variables, that is, measurements that are based on more than one item. Scales and indexes are generally used with complex variables that do not easily lend themselves to single-item or single-indicator measurement. Some items, such as newspaper circulation, can be adequately measured without scales and indexes. Others, such as attitude toward television news, generally require them.

Constructing a scale and constructing an index usually incorporate similar techniques, but there is a difference between them. Scales typically have formalized rules concerning how the multiple indicators are developed and assembled into one composite value. Indexes generally do not have detailed rules.

Perhaps the most commonly used scale in mass communication research is the Likert Scale. In this procedure, a number of statements are drawn up with respect to a topic. Respondents can either strongly agree, agree, be neutral, disagree or strongly disagree with the statements. Each response option is weighted, and each subject's responses are added to produce a single score on the topic. Figure 2.3 contains an example of a Likert Scale.

Another commonly used scaling procedure is the **semantic differential** technique. Developed by Charles Osgood and his associates (Osgood, Suci, and Tannenbaum 1957), the semantic differential is a general technique for measuring the meaning an item has for an individual. To use the technique, a name or concept is placed at the top of a series of seven-point scales anchored by bipolar attitudes. Figure 2.4 shows an example of this technique as used to measure attitudes toward *Time* magazine. The bipolar adjectives that typically "anchor" such evaluative scales are pleasant-unpleasant, valuable-worthless, honest-dishonest, nice-awful, clean-dirty fair-unfair, and good-bad. It

Figure 2.3 Example of Likert Scale items

1. (positive item) "The President should be elected by direct popular vote."

	Score Assigned
_____ strongly agree	5
_____ agree	4
_____ neutral	3
_____ disagree	2
_____ strongly disagree	1

2. (negative item) "Electing the President by direct popular vote would create more problems than it would solve."

	Score Assigned
_____ strongly agree	1
_____ agree	2
_____ neutral	3
_____ disagree	4
_____ strongly disagree	5

(Notice that the scores are reversed for a negatively worded item in order to maintain attitude measurement consistency.)

is recommended, however, that a unique set of anchoring adjectives be developed for each particular measurement situation. For example, Markham (1968), in his study of the credibility of television newscasters, used thirteen variable sets, including deep-shallow, ordered-chaotic, annoying-pleasing, and clear-hazy.

The best known index in mass communication research is probably the Violence Index, which is issued annually by Gerbner *et al.* (1980). The following measures are incorporated in this index: (1) the percentage of programs in a sample which contain violence; (2)

Figure 2.4 Example of the semantic differential technique

Time magazine

biased	_____:_____:_____:_____:_____:_____:_____	unbiased
trustworthy	_____:_____:_____:_____:_____:_____:_____	untrustworthy
valuable	_____:_____:_____:_____:_____:_____:_____	worthless
unfair	_____:_____:_____:_____:_____:_____:_____	fair
good	_____:_____:_____:_____:_____:_____:_____	bad

the frequency of violent acts per program; (3) the frequency of violent acts per hour; (4) the percentage of characters involved in violent actions; and (5) the percentage of characters involved in lethal violence.

Reliability and Validity

It is extremely important to note that developing any scale or index and using it without prior testing is poor research practice. At least one pilot study should be conducted for any newly developed scale or index to ensure that it is reliable and valid. For in order to be useful, a measurement must possess the related qualities of **reliability** and **validity.**

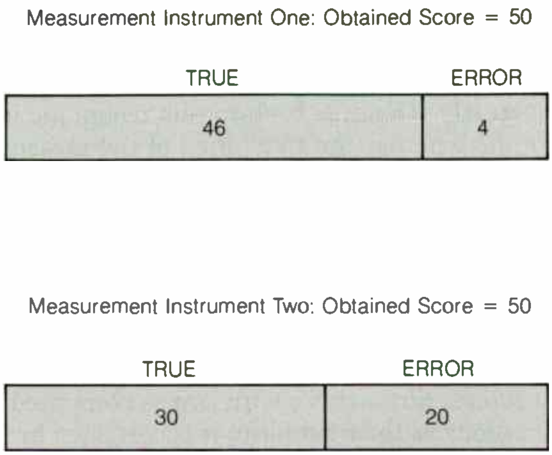
A measure is reliable if it consistently gives the same answer at different points in time. Reliability in measurement is the same as reliability in any other context. For example, a reliable person is one who is dependable, stable, and consistent. His or her behavior will be the same tomorrow as it is today. On the other hand, an unreliable person is unstable and unpredictable. He or she may act one way today and another way tomorrow. Similarly, if measurements are consistent from one session to another, they are reliable, and some degree of faith can be placed in them. If they are unreliable, it is not a good idea to depend on them.

In understanding measurement reliability, it is helpful to think of a measure as containing two components. The first represents an individual's "true" score on the measuring instrument. The second represents random error and does not provide an accurate assessment of what is being measured. Error can slip into the measurement process from several sources. Perhaps a question has been worded ambiguously, or a person's pencil slipped as he or she was filling out a measuring instrument. Whatever the cause, all measurement is subject to some degree of random error. Figure 2.5 illustrates this concept. As is evident, Measurement Instrument One is highly reliable—the ratio of the true component of the score to the total score is high. Measurement Instrument Two, however, is unreliable—the ratio of the true component to the total is low.

A completely unreliable measurement measures nothing at all. If a measure is repeatedly given to individuals and each individual's responses at a later session are unrelated to his or her earlier responses, that measure is useless. If the responses are identical or nearly identical each time the measure is given, the measure is reliable—it at least measures something (although not necessarily what the researcher intended; this problem is discussed below).

The importance of reliability should be obvious now. Unreliable measures cannot be used to detect relationships between variables. When the measurement of a variable is unreliable, it is composed mainly of random error, and random error is seldom related to anything else.

Figure 2.5 Illustration of "True" and "Error" components of a scale



An assessment of reliability is necessary in all mass media research and should be reported along with other facets of the research as an aid in interpretation and evaluation. One commonly used statistic for assessing reliability takes the form of a correlation coefficient denoted as r_{xx}. Chapter 9 provides a more detailed examination of the correlation coefficient, but for now, suffice to say that r_{xx} is a number ranging from -1.00 to $+1.00$ and is used to gauge the strength of a relationship between two variables. When r_{xx} is high—that is, approaching ± 1.00—the relationship is strong. A negative number indicates a negative relationship (high scores on one variable are associated with low scores on the other) and a positive number indicates a positive relationship (a high score goes with another high score). In measuring reliability, a high, positive r_{xx} is desired. Another common reliability coefficient is *alpha* (sometimes referred to as Cronbach's alpha), which uses the analysis of variance approach (Chapter 9) to assess the general reliability of a measure.

One method that uses correlation coefficients to determine reliability is the *test–retest method.* This procedure most closely corresponds to the conceptual definition of reliability. The same people are measured at two different points in time, and a coefficient between the two scores is computed. If r_{xx} approaches $+1.00$, it indicates that a person's score at time A was similar to his or her score at time B, showing consistency over time. There are two limitations to the test–retest technique. First, the initial administration of the measure might affect scores on the second testing. If the measuring device is a questionnaire, a person might remember his or her scores from session to session, thus falsely inflating reliability. Secondly, the concept measured may change from time A to time B, thus lowering the reliability estimate.

The *equivalent forms* technique overcomes these two problems. In this method, two equivalent forms of the same measurement instrument are administered to the same subjects, and r_{xx} is computed between the two sets of scores. The two testing sessions are placed close together in time to reduce the chance that the property being measured will change greatly. The drawback to this technique is that it is difficult to determine whether the two forms of the measure are in fact equivalent.

Another popular method of determining reliability is called the *split-half* technique. Only one administration of the measuring instrument is made, but the test is split into two halves and scored separately. For example, if the test is in the form of a questionnaire, the even numbered items might constitute one half and the odd numbered items the other. A correlation coefficient is then computed between the two sets of scores. Since this coefficient is computed from a test that is only half as long as the final form, it is corrected by using the following formula:

$$r_{xx} = \frac{2(r_{oe})}{1 + r_{oe}}$$

where r_{oe} is the correlation between the odd and the even items.

In addition to being reliable, a measurement must be valid if it is to be of use in studying variables. A valid measuring device measures what it is supposed to measure. Or, to put it in terms that were discussed earlier, determining validity requires an evaluation of the congruence between the operational definition of a variable and its conceptual or constitutive definition. Assessing validity requires some degree of judgment on the part of the researcher. In the following discussion of the major types of measurement validity, note that all of them depend, at least in part, on the judgment of the researcher. Also, validity is almost never an all-or-none proposition; it is usually a matter of degree. A measurement rarely turns out to be totally valid or invalid. Typically it winds up somewhere in the middle.

There are four major types of validity, and each has corresponding techniques for evaluating the measurement method. They are: (1) face validity; (2) predictive validity; (3) concurrent validity; and (4) construct validity.

The simplest and most basic kind of validity is *face validity.* Face validity is achieved by examining the measurement device to see if, on the face of it, it measures what it appears to measure. For example, a test designed to measure proofreading ability could include accounting problems, but this measurement would lack face validity. A test that asks people to read and correct certain paragraphs has more face validity as a measure of proofreading skill. Whether or not a measure possesses face validity depends to some degree on subjective judgment. In order to minimize subjectivity, the relevance of a given measurement should be independently judged by several individuals who are experts in that particular area.

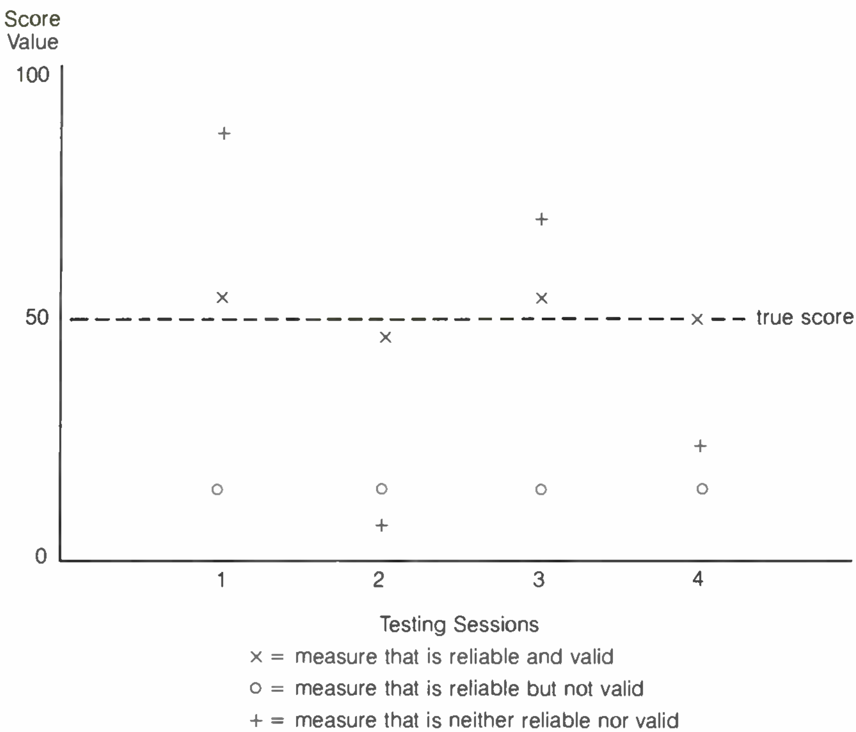
Predictive validity is characterized by checking a measurement instrument against some future outcome. For example, scores on a test to predict whether a person will vote in an upcoming election can be checked against actual voting behavior. If the test scores allow the researcher to predict with a high degree of accuracy which people will actually vote and which will not, then the test has predictive validity. Note that it is possible for a measure to have predictive validity and at the same time lack face validity. The sole factor in determining validity in the predictive method is the measurement's ability to correctly forecast future behavior. The concern is not so much what is being measured as that the measurement instrument predict something. Thus, a test to determine whether a person will become a successful mass communication researcher could conceivably consist of geometry problems. If it predicts the ultimate success of a researcher reasonably well, the test has predictive validity but little face validity. The biggest problem with predictive validity is determining the criteria against which test scores are to be checked. What, for example, constitutes a "successful mass communication researcher"? One who obtains an advanced degree? One who publishes research articles? One who writes a book?

Concurrent validity is closely related to predictive validity. In this method, however, the measuring instrument is checked against some present criterion. For example, it is possible to validate a test of proofreading ability by administering the test to a group of professional proofreaders and to a group of nonproofreaders. If the test discriminates well between the two groups, it can be said to have concurrent validity.

The last form of validity, *construct validity*, is the most complex. In simplified form, construct validity involves relating a measuring instrument to some overall theoretic framework to ensure that the measurement is actually logically related to other concepts in the framework. Ideally, a researcher should be able to suggest various kinds of relationships between the property being measured and other variables. For construct validity to exist, the researcher must show that these relationships are in fact present. For example, an investigator might expect that the frequency with which a person views a particular television newscast is influenced by his or her attitude toward that program. If the measure of attitudes correlates highly with frequency of viewing, then there is some evidence for the validity of the attitude measure. By the same token, construct validity is evidenced if the measurement instrument under consideration does *not* relate to other variables when there is no theoretic reason to expect such a relationship. In other words, if an investigator finds a relationship between a measure and other variables which are predicted by a theory and fails to find other relationships which are not predicted by a theory, then there is evidence for construct validity.

Before closing this discussion, it should be pointed out that reliability and validity are related. Reliability is necessary to establish

Figure 2.6 Relationship of reliability and validity



validity, but it is not a sufficient condition—a reliable measure is not necessarily a valid one. Figure 2.6 demonstrates this relationship.

The x's represent a test that is both reliable and valid. The scores are consistent from session to session and lie close to the "true" value. The o's represent a measure that is reliable but not valid. The scores are stable from session to session but they are not close to the true score. The +'s represent a test that is neither valid nor reliable. Scores vary widely from session to session and are not close to the true score.

Summary

Understanding empirical research requires a basic knowledge of concepts and constructs, variables, and measurement. Concepts summarize related observations and express an abstract notion that has been formed by generalizing from particulars. Connections among concepts form propositions which, in turn, are used to build theories. Constructs consist of combinations of concepts and are also useful in building theories.

Variables are things that take on one or more different values. Independent variables are those that are manipulated by the researcher; dependent variables are what the researcher attempts to explain. All variables are related to the observable world by means of operational definitions.

Measurement is a procedure in which a researcher assigns numerals to objects, events, or properties according to certain rules. There are four different levels of measurement: nominal, ordinal, interval, and ratio. To be useful, a measurement must be both reliable and valid.

Questions and Problems for Further Investigation

1. Provide operational definitions for the following items:
   a. Violence
   b. Artistic quality
   c. Programming appeal
   d. Sexual content
   e. Boredom

   Compare your definitions to those of others in the class. Would there be any difficulty in conducting a study using these definitions? Might you have demonstrated why so much controversy surrounds the topics, for example, of sex and violence on television?
2. What type of data (nominal, ordinal, interval, ratio) does each of the following concepts or measurements represent?
   a. Baseball team standings
   b. A test of listening comprehension
   c. A. C. Nielsen's list of the top ten television programs
   d. Frequency of heads vs. tails on coin flips
   e. Baseball batting averages
   f. A scale measuring intensity of attitudes toward violence
   g. VHF channels 2–13
   h. A scale for weighing yourself
3. Try to develop an unobtrusive measurement technique that would examine each of the following concepts:
   a. Newspaper reading
   b. Aggressive tendencies
   c. Brand loyalty (to purchase of products)
   d. Television viewing

References and Suggested Readings

Baran, S., Chase, L., and Courtright, J. 1979. "TV Drama as Facilitator of Prosocial Behavior." *Journal of Broadcasting* 23:3, 277–84.

Blankenburg, W. 1977. "Nixon vs. the Networks: Madison Avenue and Wall Street." *Journal of Broadcasting* 21:163–76.

Burgoon, J. and Burgoon, M. 1980. "Predictors of Newspaper Readership." *Journalism Quarterly* 57:4, 589–96.

Davidson, B. 1981. "Fact or Fiction: Television Docudramas." *Understanding Television*, ed. by R. P. Adler. New York: Praeger Publishers.

Elliot, W. and Slater, D. 1980. "Exposure, Experience and Perceived TV Reality for Adolescents." *Journalism Quarterly* 57:3, 409–14.

Elliot, W. and Schenck-Hamlin, W. 1979. "Film, Politics and the Press." *Journalism Quarterly* 56:3, 546–53.

Farace, V., and Donohew, L. 1965. "Communication in National Social Systems." *Journalism Quarterly* 42:253–61.

Gafke, R. and Leuthold, D. 1979. "A Caveat on *E&P* Poll on Newspaper Endorsements." *Journalism Quarterly* 56:2, 383–85.

Gantz, W. 1978. "How Uses and Gratifications Affect Recall of Television News." *Journalism Quarterly*, 55:4, 664–72.

Gerbner, G. *et al.* 1980. "The Mainstreaming of America: Violence Profile No. 11." *Journal of Communication* 30:3, 10–29.

Guilford, J. P. 1954. *Psychometric Methods.* New York: McGraw-Hill.

Haskins, J. and Kubas, L. 1979. "Validation of a Method for Pretesting Reader Interest in Newspaper Content." *Journalism Quarterly* 56:2, 269–76.

Kerlinger, F. 1973. *Foundations of Behavioral Research*, 2nd ed. New York: Holt, Rinehart and Winston.

Kerlinger, F. and Pedhazur, E. 1973. *Multiple Regression in Behavioral Research.* New York: Holt, Rinehart and Winston.

Levy, M. 1979. "Watching TV News as Parasocial Interaction." *Journal of Broadcasting* 23:1, 69–80.

Markham, D. 1968. "The Dimensions of Source Credibility for Television Newscasters." *Journal of Communication* 18:1, 57–64.

McNemar, Q. 1962. *Psychological Statistics.* New York: John Wiley.

Newcombe, H. 1979. *Television: The Critical View.* New York: Oxford.

Nunnally, J. 1978. *Psychometric Theory.* New York: McGraw-Hill.

Osgood, C., Suci, G., and Tannenbaum, P. 1957. *The Measurement of Meaning.* Urbana, Ill.: University of Illinois Press.

Roscoe, J. 1975. *Fundamental Research Statistics for the Behavioral Sciences.* New York: Holt, Rinehart and Winston.

Rossiter, J. and Robertson, T. 1980. "Children's Disposition Toward Proprietary Drugs and the Role of Television Drug Advertising." *Public Opinion Quarterly* 44:3, 316–29.

Thorburn, D. 1981. "Television Melodrama." *Understanding Television*, ed. by R. P. Adler. New York: Praeger Publishers.

Webb, E., Campbell, D., Schwartz, E., and Sechrest, L. 1966. *Unobtrusive Measures: Nonreactive Research in the Social Sciences.* Chicago: Rand McNally.

William, W. and Semlak, W. 1979. "Campaign, '76: Agenda Setting During the New Hampshire Primary." *Journal of Broadcasting*, 22:4, 531–39.

Wilson, C. and Howard, D. 1978. "Public Perception of Media Accuracy." *Journalism Quarterly* 55:1, 73–76.

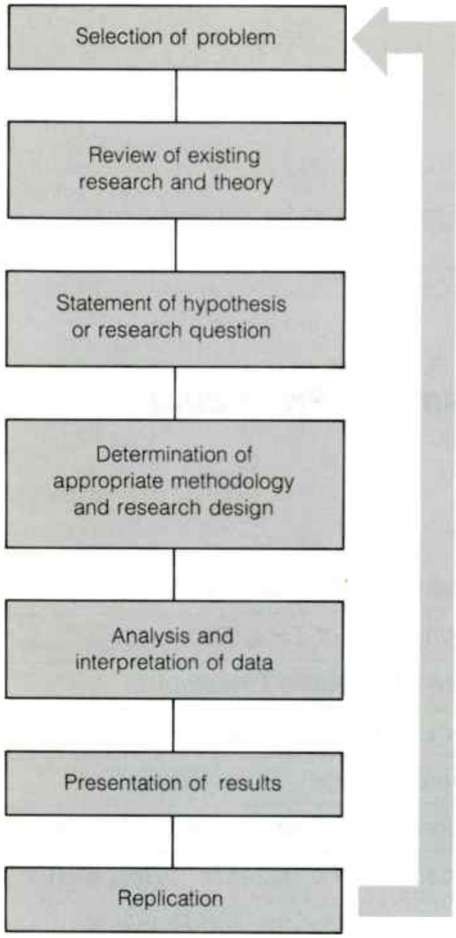
CHAPTER 3

Research Procedures

The scientific evaluation of any problem must follow a sequence of steps in order to increase the chances of producing relevant data. Researchers who do not follow a prescribed set of steps do not subscribe to the scientific method of inquiry and simply increase the amount of error present in the study. The purpose of this chapter is to describe the process involved in scientific research, from identifying and developing a topic for investigation to replication of results. The first section briefly introduces the steps involved in developing a research topic.

Objective, rigorous observation and analysis are characteristic of the scientific method. To meet this goal, researchers must follow

Figure 3.1 Steps in the development of a research project



the prescribed steps shown in Figure 3.1. This research model is appropriate to all areas of scientific research.

Selection of a Research Topic

Selecting a research topic is not a concern for all researchers; in fact, only a few investigators in communication fields are fortunate enough to be able to choose and concentrate on a research area interesting to them. Many come to be identified with specific types of studies, such as focus group methodology, magazine advertising, or communication and the law. These researchers investigate small pieces of a communication puzzle in order to obtain a broad picture of their research area.

In the private sector, researchers are often "delivered" research topics by management; they have little input into topic selection.

However, selecting a topic is a concern for many beginning researchers, especially those writing term papers, theses, and dissertations. The problem is knowing where to start. Fortunately, there are virtually unlimited sources available in searching for a research topic; academic journals, periodicals and newsweeklies, and everyday encounters can provide a wealth of ideas. Some of the primary sources are highlighted in this section.

Professional Journals

Academic communication journals, such as *Journalism Quarterly*, *Journal of Broadcasting*, and others listed below, are excellent sources for information. Although academic journals tend to publish research that is 12 to 24 months old (due to review procedures and backlog of articles), the articles may provide some ideas for research topics. Most authors conclude their research by discussing problems encountered during the study and suggesting topics that need further investigation. In addition, some journal editors use research themes for individual issues that often help in formulating research plans.

There are many high quality journals covering various aspects of research. Some journals specialize in mass media research, and others include media research at random. The following lists provide a starting point in using academic journals for research ideas.

In addition to academic journals, professional trade publica-

Journals Specializing in Mass Media Research

Journal of Broadcasting
Journalism Quarterly
Public Relations Quarterly
Journal of Marketing Research
Public Relations Review
Public Opinion Quarterly
Journal of Advertising
Journal of Marketing
Journal of Advertising Research
Journal of Consumer Research

Journals Occasionally Publishing Mass Media Research

Journal of Communication
Communication Monographs
Social Science Quarterly
Social Forces
American Psychologist
Politics
Human Communication Research
Quarterly Journal of Speech
Communication Research
Journalism Educator
Multivariate Behavioral Research
Sociology and Social Research

tions also provide a wealth of information relevant to mass media research. These include *Broadcasting, Advertising Age, Media Decisions,* and *Editor & Publisher.* The recent expansion of cable television has also created specialized publications such as *Video Review* and *Video Systems.*

Research abstracts, located in most college and university libraries, are also valuable sources for research topics. These volumes contain abstracts of research articles published in nearly every academic journal. Of particular interest to media researchers are *Communication Abstracts, Psychological Abstracts, Sociological Abstracts,* and *Dissertation Abstracts.*

Magazines and Periodicals

Many educators feel that publications other than professional journals contain only "watered-down" articles written for the general public. To some extent this is true, but these articles tend to eliminate the tedious technical jargon and are often good sources for problems and hypotheses. In addition, more and more communication articles written by highly trained professionals are appearing in weekly and monthly publications such as *TV Guide, Time,* and *Newsweek.* These sources often provide interesting perspectives on complex communication problems.

Research Summaries

Professional research organizations irregularly publish summaries that provide a close look at the major areas of research in various fields. These summaries are often useful for obtaining information about research topics, since they survey a wide variety of studies. Good examples of summary research (also known as "metaresearch") in communication include: *Television and Human Behavior,* by George Comstock *et al.; The Effects of Mass Communication on Political Behavior,* by Sydney Kraus and Dennis Davis; and *Mass Communication: A Research Bibliography,* by Donald Hansen and J. Hershel Parsons. The *Communication Yearbook,* an annual publication by the International Communication Association, has gained in popularity since its first issue in 1977 and contains summaries of the previous year's work in several communication areas. Another popular research source is the *Mass Communication Yearbook,* published by Sage Publications.

Everyday Situations

Each day we are all confronted with various types of communication via broadcasting and print, interpersonal communication, public relations campaigns, and so forth. These confrontations can be excellent sources of research topics for the researchers who take an active role in analyzing them. What types of messages are produced? Why are they produced in a specific way? What effects are expected from the various types of communication? These and other questions may help develop a research idea. Several significant studies have been based on questions arising from everyday encounters with the media and other forms of mass communication; for example, television violence, layout of newspaper advertisements, advisory warnings on television programs, and approaches to public relations campaigns.

Data Archives

Data archives, such as the Inter-University Consortium for Political Research (ICPR) at the University of Michigan, the Simmons Target Group Index (TGI), and the Gallup and Roper organizations, are another valuable source of ideas for researchers. These archives act as storage facilities where research data are deposited for use by other investigators, who conduct further investigations or ask different questions. This process, known as **secondary analysis,** has become a major research approach during the past several years because of the time and resource savings it provides.

Secondary analysis provides an opportunity for researchers to evaluate otherwise unavailable data. Becker (1981:240) defines secondary analysis as the:

> Reuse of social science data after they have been put aside by the researcher who gathered them. The reuse of the data can be by the original researcher or someone uninvolved in any way in the initial research project. The research questions examine in the secondary analysis can be related to the original research endeavor or quite distinct from it.

Advantages of Secondary Analysis. Ideally every researcher should conduct a research project of some magnitude in order to learn about design, data collection, and analysis. Unfortunately, this ideal situation does not exist. Modern research is simply too expensive. In addition, because survey methodology has become so complex, it is rare to find one researcher, or even a small group of researchers, who are experts in all phases of large studies.

Secondary analysis solves some of these problems. There is almost no expense involved in using available data. There are no questionnaires or measurement instruments to construct and validate, salaries for interviewers and other personnel are non-existent, and there are no costs for subjects and special equipment. The only expenses involved in secondary analysis are those for duplicating materials—some organizations provide their data free of charge—and computer time.

Secondary analysis has a bad connotation for some researchers, especially those who are unfamiliar with its potential (Becker 1981). Understandably, there is some value for researchers to develop questionnaires and conduct a research project using a small and oftentimes unrepresentative sample of subjects. Yet this type of analysis rarely produces externally valid results. The argument here is that in lieu of conducting a small study which has limited (if any) value to other situations, researchers would benefit from using valid data which have been previously collected.

Another advantage in secondary analysis is that data allow researchers more time to further understand what has been collected (Tukey 1969). All too often research is conducted and after a cursory analysis of the data for publication or report to management, the data are set aside never to be touched again. It is difficult to completely analyze all data from any research study in just one or two studies, yet this procedure is followed in both the academic and private sectors.

Tukey (1969:89) argues for data re-analysis especially for graduate students, but his statement applies to all researchers:

> There is merit in having a Ph.D. thesis encompass all the admitted steps of the research process. Once we recognize that research is a continuing, more or less cyclic process, however, we see that we can segment it in many places. Why should not at least a fair proportion of theses start with a reasonably careful analysis of previously collected and presumably already lightly analyzed data, a process usefully spread out over considerable time. Instant data analysis is—and will remain—an illusion.

Sociologist Herbert Hyman (1972) advocates that secondary analysis provides researchers with an opportunity to contribute to our knowledge of social processes; that such analysis allow for examination of the past.

Arguments for secondary analysis come from a variety of researchers (Hyman 1972; Tukey 1969; Glenn 1972). It is clear that the research method provides excellent opportunities to produce valuable knowledge. The procedure, however is not free from criticism.

Disadvantages of Secondary Analysis. One major flaw with secondary analysis is that researchers are limited to the hypotheses and research questions of other researchers. A person conducting a secondary analysis cannot alter the original work, and in some cases, the original hypotheses and questions might not be completely adequate. Researchers, therefore, much weigh this disadvantage to the disadvantages of gathering new data (Becker 1981).

Researchers conducting secondary analysis studies also may face the problems of using data that were poorly collected, inaccurate, or flawed in some way. Many studies do not include information about the research design, sampling procedures, weighting of subjects' responses, or other peculiarities. It might be that the data were totally fabricated. Large research firms tend to explain their procedures in detail. Although individual researchers in mass media have begun to make their data more readily available (Wimmer and Reid 1982; Reid, Soley and Wimmer 1981), not all follow adequate scientific procedures. This may seriously affect a secondary analysis.

Before selecting a secondary analysis approach, researchers need to consider the advantages and disadvantages involved. However, with the increased use of secondary analysis, some of the problems with research explanations and data storage are being solved. For an example of secondary analysis, see Becker, Bean and Russial (1978).

Determination of Topic Relevancy

Question 1: Is the Topic Too Broad?

Once a basic research idea has been chosen, the next step is to ensure that the topic has merit. This step can be accomplished by answering seven basic questions.

Most research studies concentrate on one small area of a field; few researchers attempt to analyze an entire field in one study. There is a tendency, however, for researchers to choose a topic that, while valuable, is too broad to cover in one study; for example, "the effects of television violence on children," or "the effects of mass media information on voters in a presidential election."

To avoid this problem, the researcher should first write down his or her proposed title, as a visual starting point, and then attempt to dissect the topic into small questions. Figure 3.2 illustrates this dissection process with regard to the topic "Political Communication." It is an effective way to isolate a manageable topic from a broad research category.

Question 2: Can the Problem Really Be Investigated?

Aside from considerations of broadness, a topic might prove unsuitable for investigation simply because the question being asked has no answer, or at least cannot be answered with the facilities and information available. For example, if a researcher wants to know how people who have no television react to everyday interpersonal communication situations, he or she must consider the problems of finding subjects who do not have at least one television set in the home. Some may exist in remote parts of the country, but the question is basically unanswerable due to the current saturation of television. To get around this problem, the researcher must attempt to reanalyze the original idea in conformity with practical considerations. A. S. Tan (1977) solved this particular dilemma by choosing to investigate what people do when their television sets are turned off for a period of time.

Figure 3.2 The development of a research topic from a broad category to a restricted analysis

Political Communication

Effects of mass media in presidential elections.

Effects of mass media in Senate elections.

Effects of mass media in local elections.

Mass media or interpersonal?

Local campaigns rely heavily on interpersonal communication in campaigns.

How effective is this interpersonal information on the outcome of the election?

What kinds of topics do the local candidates discuss during the campaign?

Where does the candidate get information on the various issues discussed in the campaign?

Conduct a survey of candidates to find out where they get information for speeches and interpersonal discussions.

Survey just one candidate about sources of information.

Conduct a field study: follow the candidate from the day he/she announces intention to run for office; analyze where he/she gets information for campaign issues.

Develop a workable hypothesis regarding how local political candidates develop information for issues.

He persuaded subjects not to watch television for one week and to record their use of other media, their interactions with their family and friends, and so on.

Another point to consider is whether all terms of the proposed study are definable. Remember that all measurable variables must be operationally defined (see Chapter 2). A researcher who is interested in examining youngsters' use of the media needs to come up with a working definition of the word *youngsters* in order to avoid confusion. Potential problems can be eliminated if an operational definition is stated: "Youngsters are children between the ages of three and seven."

One final consideration is to review available literature to determine whether the topic has been previously investigated. Were there any problems in previous studies? What methods were used to answer the research questions? What conclusions were drawn?

Question 3: Are the Data Susceptible to Analysis?

A topic does not lend itself to productive research if it requires the collection of data that cannot be measured reliably and validly (see

Chapter 2). In other words, the researcher who wants to measure the effects of not watching television should consider whether the information about the subjects' behavior will be adequate and reliable, whether the subjects will answer truthfully, what value the data will have once they are gathered, and so forth. Researchers also need to have enough data to make the study worthwhile. It would be inadequate to analyze only 10 subjects in the "television turn-off" example, since the results could not be generalized with regard to the entire population.

Another consideration is the researcher's previous experience with the statistical method selected to analyze the data. That is, does he or she really understand the proposed statistical analysis? Researchers need to have a basic knowledge of how the statistics work and how to interpret the results. All too often researchers design studies involving advanced statistical procedures that they have never used. This tactic invariably creates errors in computation and interpretation. Research methods and statistics should not be selected because they happen to be popular or because a research director says the method must be used, but rather because they are appropriate for a given study and are understood by the person conducting the analysis. A common error made by beginning researchers is to select a statistical method without understanding what the statistic actually produces (called the "Law of the Instrument"). It is much wiser to do simple frequencies and percentages and understand the results than to try to use a high level statistic and end up totally confused.

Question 4: Is the Problem Significant?

Before a study is conducted, the researcher must determine whether it has merit, that is, whether the results will have practical or theoretical value. The first question to ask is, "Will the results add knowledge to the information already available in the field?" The goal of all research is to help further the understanding of the problems and questions in the field of study: if a study does not do this, it has little value beyond the experience the researcher acquires from conducting it. This does not mean that all research has to be "earth-shattering." Many investigators, however, waste valuable time trying to develop monumental projects when in fact the smaller problems are of more concern.

A second question is, "What is the *real* purpose of the study?" This is important because it helps focus ideas. Is the study intended for a class paper, a thesis, a journal article, a management summary? Each of these projects has different requirements concerning background information needed, amount of explanation required, and detail of results generated. For example, applied researchers need to determine if the data will prove actionable as well as if the study will answer the question(s) posed by management.

Question 5: Can the Results of the Study Be Generalized?

For a research project to have practical value—to be significant in some way beyond the immediate analysis—it must have **external validity**; that is, one must be able to generalize from it to other situations. For example, a study of the effects of a small-town public relations campaign might be appropriate if plans are made to analyze such effects in several small towns, or if it is a case study not intended for generalization; by itself, however, the analysis has little external validity.

Question 6: What Costs and Time are Involved in the Analysis?

In many cases the cost of a research study is the sole item that determines whether the project is feasible. A researcher may have an excellent idea, but if the costs involved are prohibitive, the project must be abandoned. Private sector research is particularly expensive. Telephone surveys may cost $30 to $50 per subject, and in-person interviews may cost several hundred dollars per subject. It is common for private research firms to charge from $40,000 to $60,000 for an in-person interview study. Such prices must be considered by researchers in the planning stages of a project.

A carefully itemized list of all materials, equipment, and other facilities required is necessary before beginning a research project. If the costs seem prohibitive, the researcher must determine if he or she can shave them in some areas and still achieve the same goal. Another possibility to consider is financial aid from graduate schools, funding agencies, local governments, or other groups which provide money for research projects. In general, private sector researchers are not severely constrained by expenses; however, they must adhere to budget specifications provided by management.

Time is also an important consideration in research planning. Research studies must be designed in such a way that they can be completed in the amount of time available. Many studies have failed because not enough time was allotted for each research step, and in many cases, the pressure created by deadlines creates problems in producing reliable and valid results (for example, failure to provide attention to minor details).

Question 7: Is the Planned Approach Appropriate to the Project?

The most marvelous research idea may be greatly, and often needlessly, hindered by a poorly planned method of approach. For example, a researcher might wish to measure the change in attendance at movie theaters since television viewing has increased. The researcher knows that such information is available elsewhere, but is interested in examining only one city. He or she could mail questionnaires to a large number of people to determine how media habits have changed during the past few years. However, the costs of printing and mailing questionnaires, plus follow-up letters and possibly phone calls to increase the response rate, might prove prohibitive.

Could this study be planned differently to eliminate some of the expense? Possibly, depending on the purpose of the study and the types of questions planned. The researcher could collect the data by telephone interviews to eliminate printing and postage costs. Some questions might need reworking to fit the telephone procedure, but the essential information could be collected. A close look at every study is required to plan the best approach. Every procedure in a research study should be considered from the standpoint of the **parsimony principle,** or Occam's razor: the simplest research approach is always the most efficient.

Review of Available Literature

Researchers often spend time collecting data that are already available from another source. For example, a researcher wishing to know the amount of money spent for political advertising by the presidential candidates during a particular election campaign might send questionnaires to all the candidates in the campaign to collect the data. However, a small amount of research would reveal that this information can easily be obtained from the Federal Communications Commission. A researcher who conducts an investigation without regard to data that are already available or work that has already been done in the field is said to have fallen into the syndrome of "Ivory Tower Research." Research such as this adds nothing to our understanding of the media.

A search of the available literature saves time and money. By examining government documents, professional journals, data archives, and library sources, the researcher may find that the main portion of the data needed in a particular study has been previously collected. He or she may also discover that the chosen topic has already been the subject of one or more research studies and that a further investigation would be redundant.

In conducting a review of existing research, an investigator should bear in mind such questions as: (1) What type of research has been done in the area? (2) What has been found in previous studies? (3) What suggestions do other researchers make for further study? (4) What has not been investigated? (5) How can the proposed study add to our knowledge of the area? and (6) What research methods were used in previous studies? (Agostino 1980). Answers to these questions will usually help define a specific hypothesis or research question.

Statement of a Hypothesis

After a general research area has been identified and the existing literature reviewed, the researcher must state the problem as a workable **hypothesis** or **research question**—that is, a tentative generalization regarding the relationship between the variables that will be involved in the study. Such hypotheses are valuable because they focus a research study and suggest what data need to be gathered.

For example, Singer and Singer (1981) provide an excellent example of how a topic is narrowed, developed, and stated in simple terms. The authors were interested in whether television material enhances or inhibits a child's capacity for symbolic behavior. After a thorough review of available literature, Singer and Singer narrowed their study by seeking to answer three basic research questions:

1. Does television content *enrich* a child's imaginative capacities by offering materials and ideas for make-believe play?
2. Does television lead to *distortions* of reality for children?
3. Can *intervention* and *mediation* on the part of an adult while a child views a program, or immediately after, evoke changes in make-believe play, or stimulate make-believe play?

Research Design

The variety of research questions in mass media requires a variety of research designs. Some problems require a survey approach, via telephone or mail; others require in-person interviews. Still other problems call for a controlled laboratory situation to eliminate extraneous variables. The approach selected by the researcher depends on the goals and purpose of the study.

A research design is essentially a blueprint or set of plans for collecting information. The ideal research design collects a maximum amount of information with a minimal expenditure of time and resources. Depending on the circumstances, a design may be brief or very complicated; there are no guidelines concerning the amount of detail required.

It is important for the researcher to determine the methodology to be used in a particular project prior to beginning the study. Attempting to force a study to follow a particular approach or statistic *after* the data have been gathered only invites error. For example, a director of marketing for a large shopping mall was interested in finding out more about the customers who shopped at the mall. She designed a simple questionnaire that included questions about how often the person shopped at the mall, the person's address, and so on. The only problem was that the questionnaire was poorly designed and was not set up for easy computer analysis. She was stuck with thousands of useless questionnaires because she had not properly designed the instrument and knew nothing about computer analysis.

All procedures, including variables, samples, and questionnaire instruments, must be selected or designed in light of their appropriateness to the hypotheses or research question, and all items must be planned in advance.

There are four characteristics of research design that should be noted if a study is to produce reliable and valid results (Haskins 1968):

1. *Naturalistic setting.* In order for the results of any project to have external validity, the study must be conducted under normally encountered environmental conditions. This means that subjects should be unaware of the research situation, if possible; that phenomena should not be analyzed in a single session; and that normal intervening variables, such as noise, should be included in the study. Also, long-term projects are more conducive to a naturalistic atmosphere than short-term studies.

2. *Clear cause-and-effect relationships.* The researcher must make every effort to rule out intervening or spurious independent/dependent variable relationships. He or she can interpret the results of a study with some degree of confidence *if and only if* all confounding effects are identified.

3. *Unobtrusive and valid measurements.* There should be no perceptible connection between the communication presented to subjects and the measurement instruments used. Subjects tend to answer questions differently if they can identify the purpose of the study. Also, the study should be designed to assess both immediate and long-term effects on the subjects.

To assure the validity of the measurements used, a sample should be large enough to allow detection of minor effects or changes (see Chapter 4). Additionally, dependent variables should be selected on the basis of their relevance to the study and the researcher's knowledge of the area, not convenience.

4. *Realism.* A research design must above all be realistic. This means a careful consideration of the availability of time, money, personnel to conduct the study, and researchers who are competent in the proposed research methodology and statistical analysis.

Once the methodology is properly developed, researchers must then make a trial run, or pilot study, of the entire process. A pilot study is a small-scale version of the planned study and is designed to check for errors in the research design, measurement instrument(s), or equipment used. The mall marketing director in the previous example would have saved a great deal of time and money if she would have run a pilot study using 10 or 20 mall shoppers. She would have quickly discovered that the questionnaire did not produce the desired results. Conducting a pilot study in research is similar to taking driver's education in high school: you are allowed to make errors without jeopardizing the final results (or the real situation).

Data Analysis and Interpretation

The time and effort required for data analysis depends on the study's purpose and the methodology used. Each analysis should be carefully planned and performed according to guidelines for that analysis. Once the computations have been completed, the researchers must "step back" and consider what has been discovered. The results must be analyzed with reference to their external validity and the likelihood of their accuracy. For example, Singer and Singer (1981) summarized their findings and concluded, in part, that:

> Television by its very nature is a medium that emphasizes those very elements that are generally found in imagination: visual fluidity, time and space flexibility and make-believe .... Very little effort has emerged from producers or educators to develop age-specific programming ... it is evident that more research for the development of programming and adult mediation is urgently needed.

Researchers must determine through analysis whether their work is valid in two respects: internally and externally. This chapter has already touched briefly on the concept of external validity; an externally valid study is one whose results can be generalized to the population. **Internal validity,** on the other hand, asks: Does the study really measure or investigate the proposed research question?

Internal Validity

Control over research conditions is necessary to enable researchers to rule out all plausible rival explanations of results. Researchers are interested in verifying that "y is a function of x," or $y = f(x)$. Control over the research conditions is necessary in order to eliminate the possibility of finding that $y = f(b)$, where b is an extraneous variable. Any such variable that creates a rival explanation of results is known as an **artifact.** The presence of an artifact indicates a lack of internal validity: the study has failed to investigate its hypothesis.

Suppose, for example, that researchers discover through a study that children who view television for extended lengths of time have lower grade point averages in school than children who watch only a limited amount of television. Could an artifact have created this finding? It may be that children who view fewer hours of television also receive parental help with their school work: parental help (the artifact), and not hours of television viewed, may be the reason why grade point averages differ between the two groups.

There are several sources from which artifacts may arise. Ten of the most frequent are described below. Researchers should be familiar with these sources in order to achieve internal validity in the experiments they conduct (Campbell and Stanley 1963).

1. *History.* Various events occurring during a study may affect the subjects' attitudes, opinions, and behavior. For example, to analyze an oil company's public relations campaign for a new product, researchers first *pretest* subjects concerning their attitudes toward the company. The subjects are then exposed to an experimental promotional campaign (the *experimental treatment*); then a *posttest* is administered to determine if changes in attitude occurred as a result of the campaign. Suppose the results indicate that the public relations campaign was a complete failure—that the subjects displayed a very poor perception of the oil company in the posttest. Before the results are reported, the researchers need to determine if an intervening variable could have caused the poor perception. An investigation discloses that during the period between tests, subjects learned from a television news story that the oil company was planning to raise gasoline prices by 20%. The news of the price increase—not the public relations campaign—may have acted as an artifact that created the poor perception. The longer the time period between a pretest and a posttest, the greater the possibility that history might confound the study.

2. *Maturation.* Subjects' biological and psychological characteristics change during the course of a study. Growing hungry or tired or becoming older may influence the manner in which subjects respond to a research study. For example, fatigue will cause subjects who are analyzing three hours of television commercials to respond differently from the first commercial to the last.

3. *Testing.* Testing in itself may be an artifact, particularly when subjects are given similar pretests and posttests. A pretest may sensitize subjects to the material and improve their posttest scores regardless of the type of experimental treatment given to subjects. This is especially true when the same test is used for both situations. Subjects learn how to answer questions and to anticipate researchers' demands. To guard against the effects of testing, different pretests and posttests are required. Or, instead of being given a pretest, subjects can be tested for similarity (homogeneity) by means of a variable or set of variables that differs from the experimental variable. The pretest is not the only way to establish a *point of prior equivalency* (the groups were equal prior to the experiment) between groups—this can also be accomplished through sampling.

4. *Instrumentation.* Also known as **instrument decay,** this term refers to the deterioration of research instruments or methods which may occur over the course of a study. Equipment may wear out, observers may become more casual in their observations, and interviewers who memorize frequently asked questions may fail to present them in their proper order.

5. *Statistical regression.* Subjects who achieve either very high or very low scores on a test tend to regress to the sample or population mean during following testing sessions. Often *outliers* (subjects whose pretest scores are far from the mean) are selected for further testing or evaluation. Suppose, for example, that researchers develop a series of television programs designed to teach simple mathematical concepts, and they select only those subjects who score very low on a mathematical aptitude pretest. An experimental treatment is designed to expose these subjects to the new television series, and a posttest is given to determine if the programs increased the subjects' knowledge of the simple math concepts. The experimental study may show that indeed, after only one or two exposures to the new programs, math scores increased. But the higher scores on the posttest may not be due to the television programs; they may be a function of statistical regression. That is, regardless of whether the subjects viewed the programs, their scores may have increased merely because of statistical regression to the mean. The programs should be tested with a variety of subjects, not just those who score low on a pretest.

In addition, a **control group** (see Chapter 5) that receives no exposure to the program can help rule out statistical regression as the cause of higher posttest scores.

6. *Experimental mortality.* All research studies face the possibility that subjects will drop out for one reason or another. This is especially true in long-term studies. Subjects may become ill, move away from the town where the study is being conducted, drop out of school, or quit work. This **mortality,** or loss of subjects, is sure to have some effect on the results of the study, since most research methods and statistical analyses make assumptions about the number of subjects used. It is always better, as mentioned in Chapter 4, to select more subjects than are actually required—within the budget limits of the study.

7. *Selection.* Most research designs compare two or more groups of subjects to determine if differences exist on the dependent measurement. These groups must be randomly selected and tested for homogeneity in order to ensure that results are not due to the type of sample used.

8. *Demand characteristics.* The term **demand characteristics** is used to describe subjects' reactions to experimental conditions. Orne (1969) suggested that under some circumstances, subjects' awareness of the experimental purpose may be the sole determinant of how they behave; that is, subjects who recognize the purpose of a study may produce only "good" data for researchers.

For instance, research about television viewing habits often produces subjects who report high levels of public television viewing. However, when these same subjects are asked to list their favorite PBS programs, many cannot recall even a single one. **Cross-validating** questions—that is, the same question asked in a variety of ways—are often necessary to verify subjects' responses. In addition, researchers can help control demand characteristics by disguising the real purpose of the study; however, researchers should use caution when employing this technique (see Chapter 17).

9. *Experimenter bias.* Rosenthal (1969) discussed a variety of ways in which a researcher may influence the results of a study. Bias can enter through mistakes made in observation, data recording, mathematical computations, and interpretation. Whether experimenter errors are intentional or unintentional, they usually support the researcher's hypothesis and are considered bias (Walizer and Wienir 1978).

There are several procedures that can help to reduce experimenter bias. For example, individuals who provide instructions to subjects and make observations should not be informed of the purpose of the study; experimenters and others involved in the research should not know whether subjects belong to the experimental group or the control group (this is called a **double blind experiment**); and automated devices such as tape recorders should be used whenever possible to provide uniform instructions to subjects.

10. *Evaluation apprehension.* Rosenberg's (1965) concept of **evaluation apprehension** is similar to demand characteristics, except that subjects are essentially *afraid* of being measured or tested. They are interested in receiving only positive evaluations from the researcher and from the other subjects involved in the study. Most people are hesitant to exhibit behavior that differs from the norm and will tend to follow the group, even though they may totally disagree with them. The task of the researcher is to try to eliminate this passiveness by letting subjects know that their individual responses are important, not their unanimity as a group.

Artifacts are complex and may arise in all phases of research. For this reason, it is easy to see why the results from a single study cannot be used to refute or support a theory or hypothesis. As Hyman (1954) recognized:

> All scientific inquiry is subject to error, and it is far better to be aware of this, to study the sources in an attempt to reduce it, and to estimate the magnitude of such errors in our findings, than to be ignorant of the errors concealed in our data.

External Validity

External validity refers to the generalizability of the results of a study across populations, settings, and time (Cook and Campbell 1979). The external validity of a study can be severely affected by the interaction in an analysis of variables such as subject selection, instrumentation, and experimental conditions (Campbell and Stanley 1963).

Cook and Campbell (1979) identified three procedures that can be used to increase the external validity of a research study: (1) use random samples that are representative in some way of the population under investigation; (2) use heterogeneous samples and replicate the study several times; and (3) select a sample representative of the group to which the results will be generalized. Another way to increase external validity is to conduct research over a long period of time. Mass media research is often designed as a short-term project: subjects are exposed to an experimental treatment and are immediately tested or measured. However, in many cases, the immediate effects of a treatment are negligible. In advertising, for example, research studies designed to measure brand awareness are generally based on only one exposure to a commercial or advertisement. It is well known that persuasion and attitude change rarely take place after only one exposure; they require multiple exposures over time. Logically, such measurements should be made over a period of weeks or months to take into account the **sleeper effect:** that attitude change may be minimal or nonexistent in the short run and still prove significant in the long run.

Presentation of Results

The format used in presenting results depends on the purpose of the study. Research intended for publication in academic journals follows a format prescribed by each journal; research conducted for management in the private sector tends to be reported in simpler terms, excluding detailed explanations of sampling, methodology, and review of literature. However, all presentations of results need to be written in a clear and concise manner appropriate to both the research question and the individuals who will read the report. A more detailed discussion of reporting is included in Chapter 17.

Replication

One important point mentioned throughout this book is that the results of any single study are, by themselves, only *indications* of what might exist. A study provides information which says, in effect, "this is what may be the case." In order to be relatively certain of the results of any study, the research must be replicated. Too often, researchers conduct one study and report the results as if they are providing the basis for a theory or law. The information presented in this chapter, and in other chapters that deal with internal and external validity, argues that this cannot be true.

A research question or hypothesis requires investigation from many different perspectives before any significance can be attributed to the results of any one study. Research methods and designs must be altered to eliminate **design-specific results,** that is, results that are based on, and hence specific to, the design used. Similarly, subjects with a variety of characteristics should be studied from many angles to eliminate **sample-specific results**; and statistical analyses need variation to eliminate **method-specific results.** In other words, all effort must be made to ensure that the results of any single study are not created by or dependent upon a methodological factor; studies must be replicated.

Researchers overwhelmingly advocate the use of replication to establish scientific fact. Lykken (1968) and Kelly, Chase, and Tucker (1979) have identified four basic types of replication which can be used to help validate a scientific test.

Literal replication involves the exact duplication of a previous analysis, including the sampling procedures, experimental conditions, measuring techniques, and methods of data analysis.

Operational replication attempts to duplicate only the sampling and experimental procedures of a previous analysis in order to test whether the procedures will produce similar results.

Instrumental replication attempts to duplicate the dependent measures used in a previous study and to vary the experimental conditions under which the original study was conducted.

Constructive replication deliberately avoids imitation of a previous study in order to test the validity of the methods used; it involves varying both the manipulations and the measures used in the previous study. The researcher simply begins with a statement of empirical "fact" uncovered in a previous study and attempts to find the same "fact."

Although the process of replication has not been widely used in communication research, the trend seems to indicate that more and more mass media researchers consider it an invaluable step in producing scientific data (Wimmer and Reid 1982).

The Hazards of Research

All researchers quickly discover that research projects do not always turn out the way they were planned. It seems that Murphy's Law—"If something can go wrong, it will"—holds true in any type of research. It is therefore necessary to be prepared for difficulties, however minor, in conducting a research project. Planning and flexibility are essential. Presented below are what are known as the TAT (They're Always There) Laws. Although these "laws" are somewhat tongue-in-cheek, they are nonetheless representative of the problems one may expect to encounter in research studies.

1. A research project always takes longer than planned.
2. No matter how many people review a research proposal and say that it's perfect before you start, people will always have suggestions to make it better after the study is completed.
3. There are always errors in computer cards and on video display terminals.
4. The data errors which take the longest to find and correct are the most obvious.

5. Regardless of the amount of money requested for a research grant, the final project always costs more.
6. A computer program never runs the first time.
7. A sample is always too small.
8. Regardless of how many times a pilot study is conducted to make sure that measurement instructions are clear, there will always be at least one subject who doesn't understand the directions.
9. All electronic equipment breaks down during the most crucial part of an experiment.
10. Subjects never tell you how they really feel or what they really think or do.

Summary

The purpose of this chapter has been to describe the processes involved in identifying and developing a topic for research investigation. It was suggested that researchers consider several sources for potential ideas, including a critical analysis of everyday situations. The steps involved in developing a topic for investigation naturally become easier with experience; beginning researchers need to pay particular attention to material already available. They should not attempt to tackle broad research questions, but should try to isolate a smaller, more practical subtopic for study. They should develop an appropriate method of analysis and then proceed, through data analysis and interpretation, to a clear and concise presentation of results.

An important point stressed in the chapter is that the results of a single survey or other research approach only provide indications of what may or may not exist. Before researchers can claim support for a research question or hypothesis, the study must be replicated a number of times to eliminate a dependency on extraneous factors.

While conducting research studies, investigators must be constantly aware of potential sources of error that may create spurious results. Phenomena that effect an experiment in this way are called sources of internal validity. If and only if differing and rival hypotheses are ruled out can researchers say that it is feasible that the treatment was influential in creating differences between the experimental and control groups. A good explanation of research results rules out intervening variables; every plausible rival explanation should be considered. However, even when this is accomplished, the results of one study can only be considered as indications of what may or may not exist. Support for a theory or hypothesis can be made only after several studies are conducted that produce similar results.

In addition, for a study to have substantive worth to the understanding of mass media, the results must be generalizable to subjects and groups other than those involved in the experiment. External validity can be best achieved through randomization of subject selection: there is no substitute for random sampling.

Questions and Problems for Further Investigation

1. The focus of this chapter is on developing a research topic by first defining a major problem area and then narrowing the topic to a manageable study. Follow the procedure explained in the chapter to develop two different research projects in an area of mass media research. Use either an outline or a flow chart format.
2. Replication has long been a topic of debate in scientific research, but until recently, mass media researchers have not paid it a great deal of attention. Read the articles by Reid, Soley, and Wimmer (1981) and Wimmer and Reid (1982). Explain in your own words why replication has not been a major factor in mass media research. What could be done to correct the current situation in replication?
3. In an analysis of the effects of television viewing, it was found that the fewer the hours of television students watched per week, the higher were the scores they achieved in school. What alternative explanations or artifacts might explain the differences? How could these variables be controlled?

References and Suggested Readings

Agostino, D. 1980. "Cable Television's Impact on the Audience of Public Television." *Journal of Broadcasting* 24:3, 347–366.

Babbie, E. R. 1979. *The Practice of Social Research.* Belmont: Wadsworth.

Becker, L. B. 1981. "Secondary Analysis." *Research Methods in Mass Communication*, ed. by G. H. Stempel and B. H. Westley. Englewood Cliffs: Prentice-Hall.

Becker, L. B., Beam, R., and Russial, J. 1978. "Correlates of Daily Newspaper Performance in New England." *Journalism Quarterly* 55:100–108.

Campbell, D. T., and Stanley, J. C. 1963. *Experimental and Quasi-Experimental Designs for Research.* Chicago: Rand McNally.

Cohen, J. 1965. "Some Statistical Issues in Psychological Research." *Handbook of Clinical Psychology*, ed. by B. B. Wolman. New York: McGraw-Hill.

Comstock, G., Chaffee, S., Katzman, N., McCombs, M., and Roberts, D. 1978. *Television and Human Behavior.* New York: Columbia University Press.

Cook, T. D., and Campbell, D. T. 1979. *Quasi-Experimentation: Designs and Analysis for Field Studies.* Chicago: Rand McNally.

Glenn, N. 1972. "Archival Data on Political Attitudes: Opportunities and Pitfalls." *Political Attitudes and Public Opinion*, ed. by D. Nimmo and C. Bonjean. New York: David McKay.

Haskins, J. 1968. *How to Evaluate Mass Communication.* Chicago: Advertising Research Foundation.

Hyman, H. H. 1954. *Interviewing in Social Research.* Chicago: University of Chicago Press.

Hyman, H. H. 1972. *Secondary Analysis of Sample Surveys.* New York: Wiley.

Kelly, C. W., Chase, L. J., and Tucker, R. K. 1979. "Replication in Experimental Communication Research: An Analysis." *Human Communication Research* 5:338–42.

Kraus, S., and Davis, D. 1967. *The Effects of Mass Communication on Political Behavior.* University Park: Penn State University Press.

Lykken, D. T. 1968. "Statistical Significance in Psychological Research." *Psychological Bulletin* 21:151–59.

Orne, M. T. 1969. "Demand Characteristics and the Concept of Quasi-Controls." *Artifact in Behavioral Research,* ed. by R. Rosenthal and R. L. Rosnow. New York: Academic Press.

Reid, L. N., Soley, L. C., and Wimmer, R. D. 1981. "Replication in Advertising Research: 1977, 1978, 1979." *Journal of Advertising* 10:3–13.

Rosenberg, M. J. 1965. "When Dissonance Fails: On Eliminating Evaluation Apprehension from Attitude Measurement." *Journal of Personality and Social Psychology* 1:28–42.

Rosenthal, R. 1969. *Experimenter Effects in Behavioral Research.* New York: Appleton-Century-Crofts.

Singer, D. G., and Singer, J. L. 1981. "Television and the Developing Imagination of the Child." *Journal of Broadcasting* 25:373–87.

Tan, A. S. 1977. "Why TV Is Missed: A Functional Analysis." *Journal of Broadcasting* 21:371–80.

Tukey, J. W. 1969. "Analyzing Data: Sanctification or Detective Work?" *American Psychologist* 24:83–91.

Walizer, M. H., and Wienir, P. L. 1978. *Research Methods and Analysis: Searching for Relationships.* New York: Harper & Row.

Wimmer, R. D., and Reid, L. N. 1982. "Willingness of Communication Researchers to Respond to Replication Requests." *Journalism Quarterly* (in press).

CHAPTER 4

Sampling

This chapter describes the basics of sampling methods that are widely used in mass media research. However, considering that sampling theory has become a distinct discipline in itself, there are some studies, such as nationwide surveys, that require a consultation of more technical discussions of sampling (Raj 1972; Cochran 1963; Kish 1965; Blalock 1972).

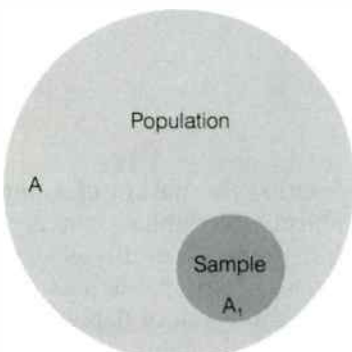
Population and Sample

One goal of scientific research is to describe the nature of a **population**, that is, a group or class of subjects, variables, concepts, or phenomena (Walizer and Wienir 1979). In some cases this is achieved through the investigation of an entire class or group, such as a study of prime time television programs during the week of September 10–16. Such a process, where every member of the population is examined, is called a **census**. In many situations, however, the chance of investigating an entire population is remote, if not nonexistent, due to time and resource constraints. The usual procedure in these instances is to select a **sample** from the population. A sample is a subset or subsegment of the population that is taken to be representative of the population. An important word in this definition is *representative*. A sample that is not representative of the population, regardless of its size, is inadequate for testing purposes: the results cannot be generalized.

The sample selection process is illustrated using a Venn diagram in Figure 4.1; the population is represented by the larger of the two spheres. A census would test or measure every element in the population (A); a sample would measure or test a segment of the population (A_1). Although Figure 4.1 might seem to indicate that the sample is drawn from only one portion of the population, it is actually selected from every portion. Assuming that a sample is chosen according to proper guidelines and is representative of the population, the results from a study using the sample can then be generalized to the population. However, generalizing results must proceed with some caution due to the inherent error involved in all sample selection methods. Theoretically, when a population is studied, only **measurement error** (that is, inconsistencies produced by the instrument used) will be present. However, when a sample is drawn from the population, the procedure introduces the likelihood of **sampling error** (that is, the degree to which measurements of the units or subjects selected differ from those of the population as a whole). Because a sample does not provide the exact data that a population would, the potential error must be taken into account.

A classic example of sampling error occurred during the 1936 presidential campaign. *Literary Digest* had predicted, based on the results of a sample survey, that the winner would be Alf Landon. Although the *Digest* sample included over a million voters, it was

Figure 4.1 The process of sample selection



composed mainly of affluent Republicans. Consequently, it was riddled with error and inaccurately represented the population. The researchers who conducted the study had failed to consider the population **parameters** (characteristics) prior to selecting their sample. As it turned out, Franklin Roosevelt was reelected president in 1936, and it may be no coincidence that the *Literary Digest* went out of business shortly thereafter.

Probability and Non-Probability Samples

A **probability sample** is selected according to mathematical guidelines whereby the chance for selection of each unit is known. A **non-probability sample** does not follow the guidelines of mathematical probability. However, the most significant distinguishing characteristic between the two types of samples is that probability sampling allows researchers to calculate the amount of sampling error present in a research study; non-probability sampling does not.

In deciding whether to use a probability or a non-probability sample, a researcher should consider four points:

1. *Cost vs. value.* The sample should produce the greatest value for the least investment. If the cost of a probability sample is too high in relation to the type and quality of information collected, then a non-probability sample is a possible alternative.

2. *Time constraints.* In many cases researchers collecting preliminary information operate under time constraints imposed by sponsoring agencies, management directives, or publication guidelines. Since probability sampling is often time-consuming, a non-probability sample may provide temporary relief.

3. *Purpose of the study.* Some research studies are not designed for generalization to the population, but rather to investigate variable relationships or to collect exploratory data for designing questionnaires or measurement instruments. A non-probability sample is often appropriate in these types of situations.

4. *Amount of error allowed.* In preliminary or pilot studies, where error control is not a prime concern, a non-probability sample is usually adequate.

Probability sampling generally incorporates some type of systematic selection procedure, such as a table of random numbers, to ensure that each unit has an equal chance of being selected. However, it does not always guarantee a representative sample from the population, even when systematic selection is followed. It is possible to randomly select 50 members of the student body at a university in order to determine the average height of all students enrolled and, by luck of the draw, end up with 50 candidates for the school's basketball team. Such an event is unlikely, but it is possible, and its possibility underscores the need to replicate any study.

Types of Non-Probability Samples

Non-probability sampling is frequently used in mass media research, particularly in the form of available samples, samples using volunteer subjects, and purposive samples. An **available sample** is a collection of readily accessible subjects for study, such as a group of students enrolled in an introductory mass media course. Although available samples can be helpful in collecting exploratory information and may produce useful data in some instances, the samples are problematic because they contain unknown quantities of error. Researchers need to consider the positive and negative qualities of available samples before using them in a research study.

Available samples are a subject of heated debate in many research fields. Critics argue that regardless of what results they may generate, available samples do not represent the population and therefore have no external validity. (This problem is discussed in Chapter 3.) Proponents of the available sample procedure claim that if a phenomenon, characteristic, or trait does in fact exist, it should exist in *any* sample. In addition, Raj (1972) contested the very notion of sample representativeness:

> Some writers suggest the use of a "representative" sample as a safeguard against the hazards of sampling. This is an undefined term which appears to convey a great deal but which is unhelpful. If it means the sample should be a miniature of the population in every respect, we do not know how to select such a sample.

Available samples can be useful in pretesting questionnaires or other preliminary (pilot study) work. They often help eliminate potential problems in research procedures, testing, and methodology before the final research study is attempted.

A sample of **volunteer subjects** is also a non-probability sample since the individuals are not selected mathematically. There is concern in all areas of research with regard to persons who willingly participate in research projects; these subjects differ greatly from non-volunteers and may consequently produce erroneous research results. Rosenthal and Rosnow (1969) identified the characteristics of volunteer subjects on the basis of several studies and found that such subjects, in comparison with non-volunteers, tend to exhibit higher educational levels, higher occupational status, greater need for approval, higher intelligence, and lower authoritarianism. They also seem to be more sociable, more "arousal-seeking," and more unconventional; they are more likely to be first children, and they are generally younger.

These characteristics mean that use of volunteer subjects may significantly bias the results of a research study and may lead to inaccurate estimates of various population parameters (Rosenthal and Rosnow 1969). Also, available data seem to indicate that volunteers may more often than non-volunteers provide data to support a researcher's hypothesis. In some cases volunteer subjects are necessary; in comparison tests of products or services, for example, these samples may be the only way the study can be conducted. However, volunteers should be used with caution because, as with available samples, there is an unknown quantity of error present in the data.

A **purposive sample** includes subjects selected on the basis of specific characteristics or qualities and eliminates those who fail to meet these criteria. Purposive samples are often used in advertising studies: researchers select subjects who use a particular type of product and ask them to compare it with a new product. A purposive sample is chosen with the knowledge that it is not representative of the general population; rather it attempts to represent a specific portion of the population. A similar method is the **quota sample** where subjects are selected in order to meet a predetermined or known percentage. For example, a researcher may be interested in finding out how VCR owners differ in their use of television than non-VCR owners, and knows that 10% of a particular population owns a VCR. The sample the researcher selected, therefore, would be comprised of 10% VCR owners and 90% non-VCR owners (to reflect the population characteristics).

Another non-probability sampling method is to select subjects haphazardly on the basis of appearance or convenience, or because they seem to meet certain requirements (the subjects *look* educated.) Haphazard selection involves researcher subjectivity and introduces error. Some haphazard samples give the illusion of a probability sample; these must be carefully approached. For example, interviewing every tenth person who walks by in a shopping center is haphazard, since not everyone in the population has an equal chance of walking by that particular location. Some people live across town, some may shop in other centers, and so on.

Types of Probability Samples

The most basic type of probability sampling is the simple **random sample,** where each subject or unit in the population has an equal chance of being selected. If a subject or unit is drawn from the population and removed from subsequent selections, the procedure is known as random sampling *without replacement*—the most widely used random sampling method. Random sampling *with replacement* involves returning the subject or unit back into the population so that it has a chance of being chosen another time. This method is often used in more complicated research studies such as nationwide surveys (Raj 1972).

Generating a simple random sample usually involves the use of a table of random numbers. To illustrate how such a table is used, suppose a researcher wants to analyze ten prime time television pro-

Table 4.1 Random Numbers

38	71	81	39	18	24	33	94	56	48	80	95	52	63	01	93	62
27	29	03	62	76	85	37	00	44	11	07	61	17	26	87	63	79
34	24	23	64	18	79	80	33	98	94	56	23	17	05	96	52	94
32	44	31	87	37	41	18	38	01	71	19	42	52	78	80	21	07
41	88	20	11	60	81	02	15	09	49	96	38	27	07	74	20	12
95	65	36	89	80	51	03	64	87	19	06	09	53	69	37	06	85
77	66	74	33	70	97	79	01	19	44	06	64	39	70	63	46	86
54	55	22	17	35	56	66	38	15	50	77	94	08	46	57	70	61
33	95	06	68	60	97	09	45	44	60	60	07	49	98	78	61	88
83	48	36	10	11	70	07	00	66	50	51	93	19	88	45	33	23
34	35	86	77	88	40	03	63	36	35	73	39	44	06	51	48	84
58	35	66	95	48	56	17	04	44	99	79	87	85	01	73	33	65
98	48	03	63	53	58	03	87	97	57	16	38	46	55	96	66	80
83	12	51	88	33	98	68	72	79	69	88	41	71	55	85	50	31
56	66	06	69	44	70	43	49	35	46	98	61	17	63	14	55	74
68	07	59	51	48	87	64	79	19	76	46	68	50	55	01	10	61
20	11	75	63	05	16	96	95	66	00	18	86	66	67	54	68	06
26	56	75	77	75	69	93	54	47	39	67	49	56	96	94	53	68
26	45	74	77	74	55	92	43	37	80	76	31	03	48	40	25	11
73	39	44	06	59	48	48	99	72	90	88	96	49	09	57	45	07
34	36	64	17	21	39	09	97	33	34	40	99	36	12	12	53	77
26	32	06	40	37	02	11	83	79	28	38	49	44	84	94	47	32
04	52	85	62	24	76	53	83	52	05	14	14	49	19	94	62	51
33	93	35	91	24	92	47	57	23	06	33	56	07	94	98	39	27
16	29	97	86	31	45	96	33	83	77	28	14	40	43	59	04	79

grams out of a total population of 100 programs to determine how these shows portray elderly people. He or she can take a random sample from the 100 programs by numbering each show from 00 to 99 and then selecting 10 numbers from a table of random numbers, such as the brief listing in Table 4.1. First, an arbitrary starting point in the table is selected. There is no specific way to choose a starting point, but it must be done randomly. The researcher then selects the remaining 9 numbers by going up, down, left, or right on the table—or even randomly throughout the table. For example, if he or she decides to go down in the table from the starting point 44 until a sample of 10 is drawn, the sample would then include those television programs numbered 44, 85, 46, 71, 17, 50, 66, 56, 03, and 49.

Simple random samples for use in telephone surveys are often obtained by a process called **random digit dialing.** One method involves randomly selecting four-digit numbers (usually generated by a computer or through the use of a random numbers table) and adding them to the three-digit exchange prefixes in the city where the survey is conducted. A single four-digit series may be used once, or it may be added to all of the prefixes.

The major problem with this method of random digit dialing is that a large number of the telephone numbers generated thereby are invalid, due to disconnected or business phones and to unused num-

Simple Random Sample

Advantages:

1. Detailed knowledge of the population is not required.
2. External validity may be statistically inferred.
3. A representative group is easily obtainable.
4. The possibility of classification error is eliminated.

Disadvantages:

1. A list of the population must be compiled.
2. A representative sample may not result in all cases.
3. Sampling error tends to be high compared to other sampling procedures.
4. The procedure can be expensive if several steps are involved.

bers. For this reason, it is advisable to produce at least three times the number of telephone numbers needed; if a sample of 100 is required, at least 300 numbers should be generated to allow for invalid numbers.

A second random digit-dialing method which tends to decrease the occurrence of invalid numbers involves adding one or two random digits to a telephone number selected from the directory. For example, if the number 546–9023 has been randomly chosen from the directory, one can use a table of random numbers to produce the "add-on." If the add-on digit is 6, the next telephone number in the sample would be 546–9029. This procedure eliminates many invalid telephone numbers because telephone companies tend to distribute numbers in series. In the example, the 90— series is already in use; this increases the probability that a phone number in the same series will be in working order. (See Fletcher 1981 for more information on random digit dialing.)

Random number generation is possible via a variety of methods, but there are always two guidelines to follow when using the procedure: (1) each unit or subject in the population must have an equal chance of being selected; and (2) the selection procedure must be free from subjective intervention by the researcher. The purpose of random sampling is to reduce sampling error; violating random sampling rules only increases the chance of introducing such error into a study.

Similar in some ways to simple random sampling is a procedure called **systematic sampling**. This involves selecting every *n*th subject or unit from a population. For example, assume that a researcher desires a sample of 20 from a population of 100, or a **sampling rate** of 1/5. He or she then randomly selects a starting point and a **sampling**

interval. Thus, if the number 11 is chosen, the sample will include the twenty subjects or items numbered 11, 22, 33, 44, and so on. To add further randomness to the process, the researcher may randomly select both the starting point and the interval. For example, an interval of 11 together with a starting point of 29 would generate the numbers 40, 51, 62, 73, and so on.

Systematic samples are frequently used in mass media research. They often save time, resources, and effort when compared to simple random samples. In fact, since the procedure so closely resembles a simple random sample, many researchers consider systematic sampling equal to the random procedure (Babbie 1979). The method is widely used in selecting subjects from lists such as telephone directories, *Broadcasting Yearbook*, and *Editor & Publisher*.

The degree of accuracy of systematic sampling depends on the adequacy of the **sampling frame,** or a complete list of members in the population. Telephone directories are inadequate sampling frames in most cases, since not all phone numbers are included, and not everyone has a telephone. However, lists that include all the members of a population have a high degree of precision. Before deciding to use systematic sampling, one should consider the goals and purpose of a study as well as the availability of a comprehensive list of the population. If such a list is not available, then systematic sampling is probably ill-advised.

One major problem with systematic sampling is that the procedure is susceptible to **periodicity;** that is, the arrangement or order of the items in the population list may bias the selection process. For example, consider the problem mentioned earlier of analyzing television programs to determine how the elderly are portrayed. Quite possibly, every tenth program listed may have been aired by ABC; the result would be a nonrepresentative sampling of the three networks. If periodicity is eliminated, systematic sampling can be an excellent sampling methodology.

A **stratified sample** is used when a researcher is interested in a

Systematic Sampling

Advantages:

1. Selection is easy.
2. Selection can be more accurate than in a simple random sample.
3. The procedure is generally inexpensive.

Disadvantages:

1. A complete list of the population must be obtained.
2. Periodicity can occur.

particular characteristic, segment, or strata of the population. Instead of selecting a sample from the population at large, the researcher identifies a significant variable and selects subjects who have this trait. He or she then chooses a subsample from this group. The variable of interest might be age, sex, religion, education, income, or political affiliation. However, the more variables that are added to the stratification list, the harder it becomes to identify subjects meeting the criteria. The term **incidence** is used to describe the frequency with which desired subjects can be found in the population. When several stratified variables are involved, the rate of incidence is low, and the time and expenses required for recruiting the sample increase.

Stratified sampling ensures that a sample is drawn from a **homogeneous** subset of the population, that is, from a population with similar characteristics. Homogeneity helps researchers to reduce sampling error. For example, consider a research study on subjects' attitudes toward two-way, interactive cable television. The investigator, knowing that cable subscribers tend to have higher achievement levels, may wish to stratify the population according to education. Before randomly selecting subjects, the researcher divides the population into three levels: grade school, high school, and college. He or she then selects a random sample proportional to the population: for example, if 10% of the population completed college, then the sample should also contain 10% who meet this standard. As Babbie (1979) noted:

> The effect of stratification is to ensure the proper representation of the stratification variables to enhance representation of other variables related to them. Taken as a whole, then, a stratified sample is likely to be more representative on a number of variables than would be the case for a simple random sample.

The usual sampling procedure is to select one unit or subject at a time. But this requires the researcher to have a complete list of the population. In some cases there is no way to obtain such a list. One way to avoid this problem is to select the sample in groups or categories; this procedure is known as **cluster sampling.** For example, analyzing magazine readership habits of people in the state of Wisconsin is time-consuming and complicated if individual subjects are randomly selected. With cluster sampling, one can divide the state into districts, counties, or zip code areas and select groups of people from these areas.

Cluster sampling creates two types of error: aside from the error involved in defining the initial clusters, errors may arise in selecting from the clusters. For example, a zip code area may comprise mostly residents of a particularly low socioeconomic status who are unrepresentative of the remainder of the state; if selected for analysis, such a group may confound the research results. To help control such

Stratified Sampling

Advantages:

1. Representativeness of relevant variables is ensured.
2. Comparisons can be made to other populations.
3. Selection is made from a homogeneous group.
4. Sampling error is reduced.

Disadvantages:

1. A knowledge of the population prior to selection is required.
2. The procedure can be costly and time-consuming.
3. It can be difficult to find a sample if incidence is low.

error, it is best to use small areas or clusters in order to decrease the number of elements in each cluster and to maximize the number of clusters selected (Babbie 1979).

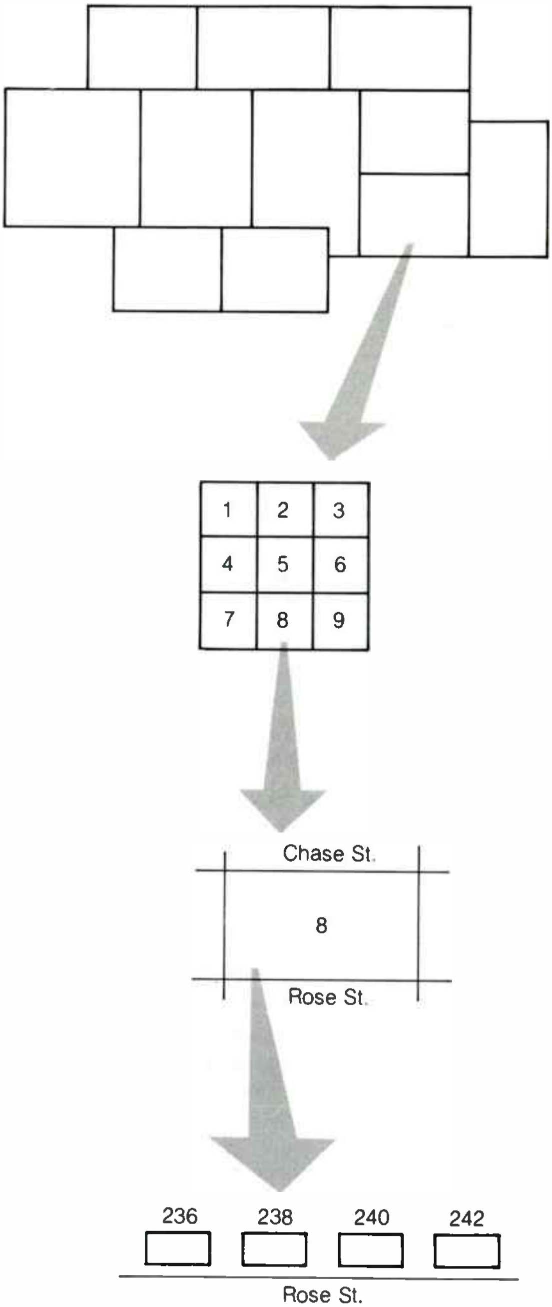
Some broadcast ratings services (see Chapter 14) use a form of cluster sampling called a *multistage sampling,* whereby several steps are involved in selecting the sample. The difference in this procedure is that individual households or persons are selected, not groups. This procedure, illustrated in Figure 4.2, involves a four-stage sequence. First, a cluster of counties in the United States is selected. This cluster is then narrowed by selecting districts or block groups within the counties, then individual blocks within each district, and finally individual households from each block. This is the method used by A. C. Nielsen.

Since the media are complex systems, researchers frequently encounter complicated sampling methods. These are known as *hybrid* situations. Consider a study where researchers are attempting to determine the potential for videotext distribution of a local newspaper to cable subscribers. This problem requires investigating readers and non-readers of the newspaper as well as cable subscribers and non-subscribers. The research, therefore, requires random sampling from the following four groups:

Group A	Cable/Newspaper Readers
Group B	Cable/Non-newspaper Readers
Group C	Non-cable/Newspaper Readers
Group D	Non-cable/Non-newspaper Readers

The researchers must identify each subject as belonging to one of these four groups. If three variables were involved, sampling from eight groups would be required, and so on. In other words, researchers are often faced with situations where sampling becomes very complicated and involves numerous steps.

Figure 4.2 Census tracts



Cluster Sampling

Advantages:

1. Only part of the population need be enumerated.
2. Costs are reduced if clusters are well-defined.
3. Estimates of cluster parameters are made and compared to the population.

Disadvantages:

1. Sampling errors are likely.
2. Clusters may not be representative of the population.
3. Each subject or unit must be assigned to a specific cluster.

Sample Size

Detemining an adequate sample size is one of the most controversial aspects of sampling. How large should a sample be in order to be representative of the population, or to provide a certain level of confidence in the results? Unfortunately, there is no simple answer. There are suggested sample sizes for various statistical procedures, but no single sample size formula or method is available for every research method or statistical procedure. For this reason, it is advisable to consult sampling texts for information concerning specific techniques (Raj 1972; Cochran 1963).

There are a few general principles to guide researchers in determining an acceptable sample size. The principles are not based on mathematical or statistical theory, but they should provide a starting point in most cases.

1. A primary consideration in determining sample size is the methodology to be used. In some cases, such as focus groups (see Chapter 6), a sample of 6 to 12 subjects is adequate if they are representative of the population under study. Small samples are also adequate for pretesting measurement instruments; for pilot studies, to determine the practicality of a research project; and for studies designed for heuristic value.

2. Sample size is almost invariably controlled by cost. Although researchers may wish to use a sample of 1,000 for a survey, the economics of such a sample are usually restrictive. Private sector research using 1,000 subjects may cost more than $100,000. Research at any level is very expensive, and these costs have great control over a project. The general rule is to use as large a sample as possible within the economic constraints of the study. If a small sample is forced upon a researcher, the results must be interpreted accordingly—that is, with caution regarding generalization.

3. Multivariate studies (see Chapter 12) always require larger samples than univariate studies because they involve the analysis of multiple response data (several measurements on the same subject). One guideline recommended for multivariate studies is: 50 = very poor; 100 = poor; 200 = fair; 300 = good; 500 = very good; 1,000 = excellent (Comrey 1973). Other researchers suggest using a sample of 100 *plus* 1 subject for each dependent variable in the analysis (Gorsuch 1974).

4. Researchers should always select a larger sample than is actually required for a study, since attrition is always a problem. Subjects drop out of research studies for one reason or another, and allowances must be made for this in planning the sample selection. Subject attrition is especially prevalent in panel studies, where the same group of subjects is tested or measured frequently over a long period of time. In most cases, researchers can expect from 10% to 25% of the sample to drop out of a study before it is completed.

5. Information about sample size is available in published research. Consulting the work of other researchers provides a base from which to start. If a survey is planned and similar research indicates that a representative sample of 400 has been used regularly with reliable results, then a sample larger than 400 may be unnecessary.

6. Generally speaking, the larger the sample used, the better. However, a large unrepresentative sample is as meaningless as a small unrepresentative sample, so researchers should not consider numbers along. Quality is always more important in sample selection than mere size.

Sampling Error

All research involves some degree of error. Research error is an additive process. That is, total error comprises sampling error plus measurement error plus **random** (unknown or uncontrollable) **error.**

Sampling error refers specifically to the sampling procedure. That is, measurements obtained from a sample differ to some degree from the measurements that would be obtained from the population. There are several ways to compute sampling error, but no single method is appropriate for all types of samples or situations. In addition, error formulas vary in complexity. One error formula, designed for estimating audience sizes during certain time periods or for certain programs and for measuring cumulative audiences (see Chapter 14), uses the standard error of a percentage derived from a simple random sample. If the sample percent is designated as p, the size of the sample as n, and the estimated or **standard error** of the sample percentage as $SE(p)$, then the formula is:

$$SE(p) = \sqrt{\frac{p(100 - p)}{n}}$$

Suppose a random sample of 500 households produces a **rating** (or estimate of the percentage of viewers; see Chapter 14) of 20 for a particular show. This means that 20% of those households were tuned in to that channel at that time. The formula can be used to calculate the standard error as follows:

$$\begin{aligned} SE(p) &= \sqrt{\frac{(20 \times 80)}{500}} \\ &= \sqrt{\frac{1{,}600}{500}} \\ &= \sqrt{3.2} \\ &= \pm 1.78 \end{aligned}$$

That is, the rating of 20 computed in the survey is subject to an error of ±1.78 points; the actual rating could be as low as 18.22 or as high as 21.78.

Standard error is directly related to sample size. The error figure improves as the sample size is increased, but in decreasing increments. Thus, an increase in sample size does not provide a quantum leap in precision. The following table shows the error rate for various sample sizes using a rating of 20:

Sample Size	*Error*	*Lower Limit*	*Upper Limit*
600	±1.63	18.37	21.63
700	±1.51	18.49	21.51
800	±1.41	18.59	21.41
900	±1.33	18.67	21.33
1,000	±1.26	18.74	21.26
1,500	±1.03	18.97	21.03

As can be seen, even with a sample of 1,500, the standard error is only .75 better than with a sample of 500. A researcher would need to determine whether the increase in time and expense caused by an additional 1,000 subjects would justify such a proportionally small increase in precision.

There are two important terms related to sampling error: **confidence interval** and **confidence level.** The confidence level, such as 95%, indicates that 95 times out of 100, a particular result will occur. A confidence interval (such as ±1.0) relates to the amount a particular value found may vary. Table 4.2 shows the amount of error at the 95% confidence level for measurements which contain dichotomous variables (such as "yes-no"). For example, with a sample of 1,000 and a 30 percent "yes" response to a question, the probable error due to sample size alone is ±2.9 percent. This means that we are 95% sure that our values for this particular question fall between 27.1% and 32.9%.

Table 4.2 Sampling error

Survey result is:		1% or 99%	5% or 95%	10% or 90%	15% or 85%	20% or 80%	25% or 75%	30% or 70%	35% or 65%	40% or 60%	45% or 55%	50%
Sample of:	25	4.0	8.7	12.0	14.3	16.0	17.3	18.3	19.1	19.6	19.8	20.0
	50	2.8	6.2	8.5	10.1	11.4	12.3	13.0	13.5	13.9	14.1	14.2
	75	2.3	5.0	6.9	8.2	9.2	10.0	10.5	11.0	11.3	11.4	11.5
	100	2.0	4.4	6.0	7.1	8.0	8.7	9.2	9.5	9.8	9.9	10.0
	150	1.6	3.6	4.9	5.9	6.6	7.1	7.5	7.8	8.0	8.1	8.2
	200	1.4	3.1	4.3	5.1	5.7	6.1	6.5	6.8	7.0	7.0	7.1
	250	1.2	2.7	3.8	4.5	5.0	5.5	5.8	6.0	6.2	6.2	6.3
	300	1.1	2.5	3.5	4.1	4.6	5.0	5.3	5.5	5.7	5.8	5.8
	400	.99	2.2	3.0	3.6	4.0	4.3	4.6	4.8	4.9	5.0	5.0
	500	.89	2.0	2.7	3.2	3.6	3.9	4.1	4.3	4.4	4.5	4.5
	600	.81	1.8	2.5	2.9	3.3	3.6	3.8	3.9	4.0	4.1	4.1
	800	.69	1.5	2.1	2.5	2.8	3.0	3.2	3.3	3.4	3.5	3.5
	1000	.63	1.4	1.9	2.3	2.6	2.8	2.9	3.1	3.1	3.2	3.2
	2000	.44	.96	1.3	1.6	1.8	1.9	2.0	2.1	2.2	2.2	2.2
	5000	.28	.62	.85	1.0	1.1	1.2	1.3	1.4	1.4	1.4	1.4

Source: *A Broadcast Research Primer*. Washington, D.C.: National Association of Broadcasters, 1976, p. 19.
Reprinted by permission.

Sampling error is an important concept in all research areas because it provides an indication of the degree of accuracy of the research. Research studies published by large audience measurement firms such as Arbitron and A. C. Nielsen contain simplified charts to determine sampling error (they are required to by the Broadcast Ratings Council). In addition, each company provides some type of explanation about error, such as the Arbitron statement:

> Sampling error in Arbitron estimates are subject to the statistical variances associated with all surveys using a sample of the universe. . . . the accuracy of Arbitron measurements, data and reports and their statistical evaluators cannot be determined to any precise mathematical value or definition (Description of Methodology, 1980).

Statistical error due to sampling is a characteristic of all research studies.

Researchers must pay specific attention to the potential sources of error in any study. Producing a study riddled with error is tantamount to never having conducted the study at all. The problem with error is that its magnitude can never be accurately assessed. If this were possible, researchers could simply determine the source of error and correct it. Since they cannot, however, they must accept error as part of the research process, attempt to reduce its effects as much as possible, and remember always to interpret their results with regard to its presence.

Sample Weighting

In an ideal study, a researcher will have enough subjects to represent the demographic, **psychographic** (why people behave in specific ways), and lifestyle characteristics of the population. The ideal sample, however, is rare due to the time and budget constraints of most research. Instead of cancelling a research project due to sampling inadequacies, most researchers utilize a statistical procedure known as **weighting.** That is, in categories where subject totals do not reach the population percentages, subjects' responses are multiplied (weighted) to allow for the shortfall. A single subject's responses may be multiplied by 1.3, 1.7, 2.0, or any other figure to reach the predetermined required level.

Subject weighting is a controversial data manipulation technique, especially in the area of broadcast ratings. The major question is just how much one subject's responses can be weighted and still be representative. Weighting is discussed in greater detail in Chapter 14.

Summary

In order to make predictions about events, concepts, or phenomena, researchers must perform detailed, objective analyses. One procedure to use in such analyses is a census, where every member of the population is studied. Conducting a census for every research project is impractical, however, and researchers must resort to alternative methods. The most widely used alternative is to select a random sample from the population, examine it, and then make predictions from it that can be generalized to the population.

If the scientific procedure is to provide valid and useful results, researchers must pay close attention to the methods they use in selecting a sample. This chapter has described several types of samples commonly used in mass media research. Some are elementary and do not require a great deal of time or resources. Other sampling methods entail great expense and time. Researchers must decide whether costs and time are justified in relation to the results generated.

Sampling procedures must not be taken lightly in the process of scientific investigation. It makes no sense to develop a research design for testing a valuable hypothesis or research question and then nullify this effort by neglecting correct sampling procedures. These procedures must be continually scrutinized to ensure that the results of an analysis are not sample-specific.

Questions and Problems for Further Investigation

1. The use of available samples in research has long been a target for heated debate. There are proponents of each side of the argument. Some researchers say that available samples are inaccurate representations of the population; others claim that if a concept or phenomenon exists, it should exist in an available sample as well as in a random sample. Which argument do you agree with? Explain your answer.
2. Many research studies use small samples. What are the advantages or disadvantages of such samples? Can anything be gained by using a small sample in a research study, other than cost savings?
3. What sampling technique might be appropriate for the following research projects?
    a. A pilot study to test whether people understand the directions to a telephone questionnaire.
    b. A study to determine who buys video cassette recorders.
    c. A study to determine the demographic makeup of the audience for a local television show.
    d. A content analysis of commercials during Saturday morning children's programs.
    e. A survey examining the differences between newspaper readership in high- and low-income households.

References and Suggested Readings

Babbie, E. R. 1979. *The Practice of Social Research.* Belmont: Wadsworth.

Blalock, H. M. 1972. *Social Statistics.* New York: McGraw-Hill.

Cochran, W. G. 1963. *Sampling Techniques.* New York: Wiley.

Comrey, A. L. 1973. *A First Course in Factor Analysis.* New York: Academic Press.

Fletcher, J. E., ed. 1981. *Handbook of Radio and TV Broadcasting.* New York: Van Nostrand Reinhold.

Gorsuch, R. L. 1974. *Factor Analysis.* Philadelphia: W. B. Saunders.

Kish, L. 1965. *Survey Sampling.* New York: Wiley.

Nunnally, J. C. 1978. *Psychometric Theory.* New York: McGraw-Hill.

Raj, D. 1972. *The Design of Sample Surveys.* New York: McGraw-Hill.

Rosenthal, R., and Rosnow, R. L. 1969. *Artifact in Behavioral Research.* New York: Academic Press.

Walizer, M. H., and Wienir, P. L. 1979. *Research Methods and Analysis.* New York: Harper & Row.

P A R T 2

Research Approaches

CHAPTER 5

Laboratory Research and Experimental Design

The mass media are complicated phenomena. Not only are there a variety of media presenting information to audience members, the audience uses these media in a variety of ways. Some people use the media for information, others use them for entertainment or just to pass time. These complicated situations create a variety of research opportunities for investigators. One research approach cannot be used because there are simply too many different situations that need investigation. Some situations require a laboratory approach where subjects are studied under controlled conditions; other situations require telephone or in-person interviews.

The purpose of this chapter is to describe the laboratory method of research and its use in mass media investigations. The laboratory method is the oldest approach in mass media research and continues to provide a wealth of information for researchers and critics of the media.

Laboratory ("lab") research has been a popular approach since the early 1900s, for at least two reasons. First, lab experiments follow a fairly simple format. They involve measuring the relationship between two types of variables under closely observed and controlled conditions. Secondly, lab experiments are often valuable in determining causation, since variable manipulation is controlled by the researcher.

The lab approach has five basic uses in mass media research:

1. To investigate hypotheses and research questions under controlled conditions
2. To develop theories which can be tested in the field
3. To refine existing theories and research findings
4. To investigate problems in small steps
5. To ease replication, since conditions of the study are clearly specified

Controls

The fact that lab research is conducted under controlled conditions is particularly important. Researchers using the lab method have control over three areas of the project: the environment, the variables, and the subjects.

Environmental control. Lab research allows investigators to isolate a testing situation from the competing influences of normal activity. Researchers are free to structure the experimental environment in almost any way. Lighting and temperature levels, proximity of subjects to measurement instruments, soundproofing, and nearly every other aspect of the experimental situation can be arranged and altered.

However, the environmental artificiality created in lab experiments has been a thorn in the side of lab research for years. Critics claim that the sterile and unnatural conditions created in the laboratory produce results that have little direct application to "real world settings," where subjects are continually exposed to competing stimuli. Miller (1970) noted that critics of the lab method often resort to ambiguous and disjunctive arguments about the artificiality of the procedure; that contrasting the "real" world with the "unreal" world may, in fact, be merely a problem in semantics. The main point he claimed, is that both lab and field methods investigate communication behavior, and if viewed in this way, it is meaningless to speak of behavior as "real" or "unreal": all behavior is real.

Miller also noted, however, that it is unsatisfactory and unscientific to dodge the problem of artificiality in lab procedures by including a disclaimer to a study indicating that the findings are applicable only to the particular audience involved, the environmental conditions of the analysis, and the given time period during which the study was conducted. Since external validity is a major goal of scientific research, a disclaimer of this nature is counterproductive. If researchers are not willing to expand their interests beyond the scope of a single analysis, then such studies have only heuristic value; they make little or no contribution to the advancement of knowledge in mass media.

Laboratory studies can be particularly valuable as an initial phase of research. Problems can be defined, methods can be tested, and the worth of a study can be judged. The results of such studies are not necessarily invalid or irrelevant. For example, suppose researchers are interested in determining how television soap operas help college students solve communication problems—a question that has not been previously investigated. Assume a preliminary lab study finds that a number of college students do use soap operas as models in their own communication encounters. Would this finding be totally irrelevant? The answer is no. The researchers might conclude their preliminary study with the following statement:

> The purpose of this study was to conduct a preliminary analysis of the use of television soap operas in solving communication problems. The results indicated that 25% of the sample feel that soap operas do in fact help them solve communication problems in their own lives. This leads to speculation that other samples, and other groups in the general population, may also share these attitudes. We suggest further research using other types of samples, different soap operas, and different research designs. It is possible that a substantial number of viewers in the general population reflect the attitudes and behaviors of the subjects uncovered in the analysis.

This explanation neither totally supports nor totally discounts the findings of the hypothetical study. It states only that an attitude or behavior is present and that these same characteristics may quite likely be found in other samples using different research procedures.

Severe criticisms of the lab method (in reference to statistical analysis) may subside when more reliable measurements have been developed. In addition, not all lab experiments involve wiring subjects to machines or placing them in soundproof rooms; many more "natural" studies, such as measurement development (see Nunnally 1978 for information on measurement development), are particularly suited to the lab approach.

Variable control. Laboratory studies also allow researchers to control the number and types of independent and dependent variables selected and the way these variables are manipulated. Variable control strengthens internal validity and helps eliminate confounding influences. Appel, Weinstein, and Weinstein (1979), for example, were able to control almost every detail of their laboratory analysis of alpha brain wave activity while viewing television commercials.

Control over subject selection. More than any other research method, the lab approach allows researchers to control subjects. This includes control over the selection process, assignment to the experimental or the control group, and exposure to the experimental treatment. Researchers can place limits on the number of subjects involved in a study as well as choose specific types of subjects for exposure to varying degrees of the independent variable. For example, they may select subjects according to which medium they use for news information and then vary each subject's exposure to different types of commercials to determine which is the most effective.

Control is important in experimentation because it allows researchers to rule out rival explanations of results. The laboratory method, although often criticized for its lack of external validity, does offer the benefits of control.

Experimental Design

Experimental design refers to the steps involved in conducting laboratory research. An experimental design does not have to be a complicated series of statements, diagrams, and figures; it may be as simple as:

Pretest ⟶ Experimental Treatment ⟶ Posttest

Although other factors, such as variable and sample selection, control, and construction of a measurement instrument, enter into this design, the diagram does provide a legitimate starting point for research.

To facilitate the discussion of experimental design, the following notations are used to represent specific parts of a design (Campbell and Stanley 1963):

R represents a random sample or random assignment.

X represents a treatment or manipulation of the independent variables so that the effects of these variables on the dependent variables can be measured.

O refers to a process of observation or measurement; it is usually followed by a numerical subscript indicating the number of the observation (O_1 = "observation 1").

A left-to-right listing of symbols, such as R O_1 X O_2, represents the order of the experiment. In this case, subjects are randomly selected or assigned to groups (R) and then observed or measured (O_1). Next, some type of treatment or manipulation of the independent variable is performed (X), followed by a second observation or measurement (O_2). Each line in experimental notation refers to the experience of a single group. A design such as:

R O_1 X O_2

R O_1 $\quad$ O_2

indicates that the operations in the experiment are conducted simultaneously on two different groups. Notice that the second group (the control group) does not receive the experimental treatment.

Basic Experimental Designs

Research designs are as unique and varied as the questions and hypotheses they help answer. Different types of designs yield different types of information. If information about the effects of time is desired, then a **repeated measures design** (several measurements on the same subjects) or a panel (long-term) design can be used. If information about the effects of experimental treatment *order* is desired, then a procedure known as a *Latin Square* design is necessary, where the order of presenting the independent variable is changed to help control error.

Each experimental design makes assumptions about the type of data the researcher wishes to collect, since different data require different research methods. Several questions need to be answered by the researcher before any type of design is constructed:

1. What is the purpose of the study?
2. What is to be measured or tested?
3. How many factors (independent variables) are involved?
4. How many levels of the factors (varying degrees of the independent variables) are involved?
5. What type of data is desired?
6. What is the easiest and most efficient way to collect the data?
7. What type of statistical analysis is appropriate for the data?
8. How much will the study cost?
9. How can these costs be trimmed?
10. What facilities are available for conducting the study?
11. What types of studies have previously been conducted in the area?
12. What benefits will be received from the results of the study?

The answer to each question has a bearing on the sequence of steps a study should follow. For example, if only a limited budget is available for the study, a complicated, four-group research design must be excluded. Or, if previous studies have shown the "posttest only" design to be useful, another design may be unjustified.

Not all experimental designs are covered in this section; only the most widely used are considered. The sources listed at the end of the chapter provide more information about these and other designs.

Pretest-Posttest Control Group. The pretest-posttest control group design is a fundamental and widely used procedure in all research areas. The design controls many of the rival hypotheses generated by artifacts: the effects of maturation, testing, history, and other sources are controlled, because each group faces the same circumstances in the study. The design takes the form shown in Figure 5.1.

Figure 5.1 Pretest-posttest control group design

R	O_1	X	O_2
R	O_1		O_2

It specifies that subjects are to be randomly selected and each group given a pretest. Only the first group, however, is to receive the experimental treatment. The difference between O_1 and O_2 for Group 1 is compared to the difference between O_1 and O_2 for Group 2. If a significant statistical difference is found, it is assumed that the experimental treatment was the primary cause.

Posttest Only Control Group. When researchers are hesitant to use a pretest because of the possibility of subject sensitization to the posttest, the design in Figure 5.1 can be altered to describe a posttest only control group, shown in Figure 5.2. No pretest is involved for either group. The design involves exposing Group 1 to the treatment variable, followed by a posttest. The two groups are then compared to determine if a statistical significance is present.

The posttest only control group design is also widely used to control rival explanations. Both groups are equally affected by maturation, history, and so on. Also, both normally call for a t-test, a test to compare the significance between two groups, to determine if a significant statistical difference is present.

Solomon Four Group Design. The Solomon Four group design combines the designs shown in Figures 5.1 and 5.2 and is useful if pretesting is considered to be a negative factor. The design notation is given in Figure 5.3. Each alternative for pretesting and posttesting is accounted for in the design, which makes it attractive to researchers. The main drawback is that four separate groups are required—a luxury many researchers cannot afford.

The Solomon Four group design, in addition to controlling extraneous variables, considers the aspect of external validity, since effects and interaction of testing are determinable (two groups receive

Figure 5.2 Posttest only control group design

R	X	O_1
R		O_2

Figure 5.3 Solomon four group design

R	O_1	X	O_2
R	O_3		O_4
R		X	O_5
R			O_6

the pretest, and two do not). It allows for greater generalizability than the designs previously discussed.

Factorial Studies

Research studies involving the simultaneous analysis of two or more independent variables are called **factorial designs,** and each independent variable is called a *factor.* The approach saves time, money, and resources and allows researchers to investigate the interaction between the independent variables. That is, in many instances, it is possible that two or more variables are *interdependent* in the effects they produce on the dependent variable, a relationship that could not be detected if two simple randomized designs were used.

The term *two factor design* indicates that two independent variables are manipulated; a three factor design includes three independent variables, and so on. (A one factor design is a simple random design, because only one independent variable is involved.) At least two factors or independent variables are required in a study before it is called a "factorial" design.

Factors may also have two or more levels. Therefore, the term *2 × 2 factorial design* means "two independent variables, each with two levels." A 3 × 3 factorial design has three levels for each of the two independent variables. A 2 × 3 × 3 factorial design has three independent variables: the first has two levels, and the second and third have three levels each.

To demonstrate the concept of levels, imagine that a television station manager would like to study the success of a promotional campaign for a new movie-of-the-week series. The manager plans to advertise the new series on radio and in the newspaper. Subjects selected randomly are placed into one of the *cells* of the 2 × 2 factorial design in Figure 5.4. In this manner she can test two levels of two independent variables, exposure to radio and exposure to newspapers.

Four groups are involved in the study: Group I receives exposure to both newspaper and radio materials; Group II is exposed only to newspaper; Group III is exposed only to the radio; and Group IV

Figure 5.4 2 × 2 factorial design

	Radio	No radio
Newspapers	I	II
No newspapers	III	IV

serves as a control group and receives no exposure to either radio or newspaper. After the groups have undergone the experimental treatment, the manager can administer to them a short questionnaire to determine which medium, or combination of media, worked most effectively.

A 2 × 3 factorial design, which adds a third level to the second independent variable, is shown in Figure 5.5. This design demonstrates how the manager might investigate the relative effectiveness of full color vs. black and white newspaper advertisements while also measuring the impact of the exposure to radio materials.

Say the television station manager wants to include promotional advertisements on television as well as using radio and newspaper. The third factor produces a 2 × 2 × 2 factorial design. This three factor design is more difficult to represent pictorially (see Figure 5.6). This design shows the eight possibilities of a 2 × 2 × 2 factorial study. Note that the subjects in Group I are exposed to newspaper, radio, and television announcements, while those in Group VIII are not exposed to any of the announcements.

The testing procedure in the three factor design is similar to previous methods. Subjects in all eight cells would be given some type

Figure 5.5 2 × 3 factorial design

	Radio	No radio
Full color newspaper ad	I	II
Black/white newspaper ad	III	IV
No newspaper	V	VI

Figure 5.6 2 × 2 × 2 factorial design

	Radio		No radio	
	TV	No TV	TV	No TV
Newspaper	I	II	III	IV
No newspaper	V	VI	VII	VIII

of measurement instrument, and differences between the groups would be tested for statistical significance.

Time Series Designs

Longitudinal or **time series designs,** whereby subjects are tested or interviewed several times over a long period of time, are becoming more popular in mass media research. They would be used a great deal more if not for the high cost of such studies and the amount of time necessary to complete an analysis.

Basically, there are two types of time series designs: **panel studies** and **trend analyses.** Both procedures are similar in that they investigate a problem over a series of intervals rather than during one experiment. The primary difference between the two approaches is that the panel study involves repeated measurements of the same sample, whereas the trend study examines different but equivalent samples. Panel designs are attractive to researchers measuring long-term media effects; trend analyses are more often used to determine changes in population characteristics, such as birth rate, family size, and opinions toward concepts or subjects. The Gallup and Roper organizations, for example, are known for their trend studies of Americans' attitudes concerning governmental and societal issues.

A time series design involving one measurement can be pictured as shown in Figure 5.7. The number of observations and manipulations of independent variables depends on the research study. A panel study investigating long-term effects of television viewing could begin with young children and follow them to age 60.

Figure 5.7 Time series design

O_1 O_2 O_3 O_4 X O_5 O_6 O_7 O_8 (experimental)

O_1 O_2 O_3 O_4 O_5 O_6 O_7 O_8 (control)

Literally hundreds of observations could be made during this time. Bishop, Sharma and Brazel (1980) have provided an example of the time series approach in their analysis of newspapers. The authors developed a time series model to account for the growth in demand for daily newspapers in the United States.

Arundale (1980) suggested that many time series designs are faulty because they fail to consider two important criteria: (1) How long must measurements continue? and (2) How frequently must measurements be made?

Laboratory Research Example

An example from published literature is used here to illustrate the various facets of research design and the laboratory method. Cantor and Venus (1980) were interested in the effects of humor on the ability to recall radio commercials. While it would have been possible to study this topic by conducting a survey or by launching a field experiment, these researchers chose to investigate the problem under tightly controlled lab conditions. The authors were dissatisfied with the ambiguity of existing literature relating to humor and advertising and decided to investigate the problem using a broad research question: they decided to test the "effect of humor on the memorability and persuasiveness of a rigorously manipulated radio advertisement which was heard in a quasi-naturalistic setting."

Step 1: Experimental Design. After reviewing the available literature in the field and defining their research question, Cantor and Venus designed a 2 × 2 × 2 factorial study, shown in Figure 5.8.

Step 2: Sample Selection and Assignment. The authors recruited undergraduate students at the University of Wisconsin (59 male, 58 female) as subjects for the experiment and assigned them randomly to the four different exposure groups for each sex.

Figure 5.8 Cantor and Venus study design

	Serious		Humorous	
	Serious intro	Humorous intro	Serious intro	Humorous intro
Male				
Female				

Step 3: Manipulation. The experiment was designed to make subjects believe they were involved in three separate studies; in reality, however, two of the experiments were fabricated to disguise the "real" experiment—listening to preproduced radio recordings. The recordings contained: (1) a serious commercial preceded by a humorous programming segment; (2) a humorous commercial preceded by a serious programming segment; (3) a serious commercial preceded by a serious programming segment; and (4) a humorous commercial preceded by a humorous programming segment. In all cases, the content of the commercial was the same; only the approach was changed. Additionally, Cantor and Venus studied the differences between male and female perceptions of the commercials.

This demonstrates how variables can be manipulated in the lab approach. The researchers controlled when, how, and where subjects would be exposed to the commercials and program segments. The messages were also controlled, since they were produced specifically for the experiment.

Step 4: Gathering Data. Cantor and Venus first escorted subjects into a cubicle, where they heard a radio message, then transferred them to another cubicle, where they were exposed to the experimental treatment. Finally, subjects were taken to another room to complete a survey and participate in an experiment of recall.

Step 5: Data Analysis. The authors selected analysis of variance (see Chapter 11) as an appropriate statistic to analyze their data. The article must be read to fully understand the details of their analysis; basically, however, they discovered that the presence or absence of humor in the commercials had no appreciable effect on attitudes toward the products advertised.

Summary

Mass media researchers have a number of research designs from which to choose when analyzing a given topic. The laboratory experiment has been a staple tool in mass media research for several decades. Although the lab experiment is criticized by many researchers as being artificial, the method offers a number of advantages that make it particularly useful to some researchers. Of specific importance is the researcher's ability to control the experimental situation and to manipulate experimental treatments.

This chapter also described the process of experimental design, the blueprint that researchers use in order to conduct an experiment. The experimental design provides the steps the researcher will follow to prove or disprove an hypothesis or research question. Some experimental designs are very simple and take very little time to perform; others involve many different groups and numerous treatments.

The laboratory research approach may have faults, as do all other research procedures, but the method has proved efficient in many instances and will continue to be used by mass media researchers.

Questions and Problems for Further Investigation

1. Provide four research questions or hypotheses for any mass media area. Which of the designs described in this chapter is best suited to investigate the problems?
2. What are the advantages and/or disadvantages of the following experimental designs?

a.			X	O_1
				O_2
b.	R		X	O_1
c.	R	O_1	X	O_2
	R		X	O_3
d.	R	O_1	X	O_2

3. The Cantor and Venus laboratory study on the effect of humor on radio advertisements was described in this chapter. In the same issue of the *Journal of Broadcasting* (Volume 24:1), Krull and Husson (pp. 35–47) investigated children's perceptions of television. How does this study compare to the laboratory methods used by Cantor and Venus?

References and Suggested Readings

Appel, V., Weinstein, S., and Weinstein, C. 1979. "Brain Activity and Recall of TV Advertising." *Journal of Advertising Research* 19:4, 7–18.

Arundale, R. B. 1980. "Studying Change Over Time: Criteria for Sampling from Continuous Variables." *Communication Research* 7:2, 227–63.

Babbie, E. R. 1979. *The Practice of Social Research.* Belmont: Wadsworth.

Bishop, R. L., Sharma, K., and Brazel, R. J. 1980. "Determinants of Newspaper Circulation: A Pooled Cross-Sectional Time-Series Study in the United States 1850–1970." *Communication Research* 7:1, 3–32.

Bruning, J. L., and Kintz, B. L. 1977. *Computational Handbook of Statistics.* Glenview: Scott, Foresman and Company.

Campbell, D. T., and Stanley, J. C. 1963. *Experimental and Quasi-Experimental Designs and Research.* Chicago: Rand McNally.

Cantor, J., and Venus, P. 1980. "The Effects of Humor on Recall of a Radio Advertisement." *Journal of Broadcasting* 24:1, 13–22.

Cook, T. D., and Campbell, D. T. 1979. *Quasi-Experimentation: Designs and Analysis for Field Studies.* Chicago: Rand McNally.

Haskins, J. B. 1968. *How to Evaluate Mass Communication.* New York: Advertising Research Monograph.

Keppel, G. 1973. *Design and Analysis: A Researcher's Handbook.* Englewood Cliffs: Prentice-Hall.

Miller, D. C. 1977. *Handbook of Research Design and Social Measurement.* 3rd ed. New York: David McKay.

Nunnally, J. C. 1978. *Psychometric Theory.* New York: McGraw-Hill.

Roscoe, J. T. 1975. *Fundamental Research Statistics for the Behavioral Sciences.* New York: Holt, Rinehart and Winston.

Rosenthal, R., and Rosnow, R. L. 1969. *Artifact in Behavioral Research.* New York: Academic Press.

Rosenthal, R. 1969. *Experimenter Effects in Behavioral Research.* New York: Appleton-Century-Crofts.

Rosenberg, M. J. 1965. "When Dissonance Fails: On Eliminating Evaluation Apprehension from Attitude Measurement." *Journal of Personality and Social Psychology* 1, 28–42.

Senter, R. J. 1969. *Analysis of Data.* Glenview: Scott, Foresman.

Walizer, M. H., and Wienir, P. L. 1978. *Research Methods and Analysis: Searching for Relationships.* New York: Harper & Row.

CHAPTER 6

Field Research

Mass media researchers use a variety of methods to investigate problems. The laboratory approach, which was discussed in Chapter 5, is not suitable for all research situations. One alternative to the laboratory method is field research. This chapter discusses four types of field research: field experiments, field observations, focus groups, and panel studies.

Field Experiments

Experiments conducted in a laboratory, as seen in Chapter 5, can be disadvantageous for many research studies because of certain problems they present: they are performed in controlled conditions that are unlike natural settings; they are generally considered to lack external validity; and they usually necessitate subject awareness of the testing situation. Because of these shortcomings, many researchers prefer to use field experiments (Haskins 1968).

The exact difference between lab and field experiments has been a subject of debate for years, especially with regard to the "realism" of the situations involved (Redding 1970). Many researchers consider field and lab experiments to be on opposite ends of the "realism" continuum. However, the main difference between the two approaches is the setting. As Westley (1981) pointed out:

> The laboratory experiment is carried out on the experimenter's own turf; the subjects come into the laboratory. In the field experiment, the experimenter goes to the subject's turf. In general, the physical controls available in the laboratory are greater than those found in the field. For that reason, statistical controls are often substituted for physical controls in the field.

The two approaches can also be distinguished by the presence or absence of rules and procedures to control the conditions and the subjects' awareness or unawareness of being subjects. If the researcher maintains tight control over the subjects' behavior and the subjects are placed in a situation they perceive to be radically different from their everyday life, the situation is probably better described as a laboratory experiment. On the other hand, if the subjects function primarily in their everyday social roles with little investigator interference or environmental restructuring, then the situation is probably closer to a field experiment. Basically, the difference between lab experiments and field experiments is one of degree (for a detailed discussion of this topic, see Redding 1970).

Advantages of Field Experiments

The major advantage of field experiments is their external validity: since study conditions closely resemble natural settings, subjects usually provide a truer picture of their normal behavior and are not influenced by the experimental situation. For example, consider a lab study designed to test the effectiveness of two different versions of a television commercial. One group views Version A and the other views Version B. Both groups are then given a questionnaire to measure their willingness to purchase the advertised product. On the basis of these results, it may be concluded that Version B is more effective in selling the product. While this may actually be the case, the validity of the experiment is questionable because the subjects were aware they were being studied; they may have thought, for example, that the commercial was terrible, but reported that they would purchase the product in order to please the researcher (see the discussion of demand characteristics in Chapter 3). Another problem is that answering a questionnaire cannot be equated to actually buying the product. Furthermore, viewing commercials in a laboratory setting is different from the normal viewing situation, in which competing stimuli (crying children, ringing telephones, and so on) are often present.

In a field experiment, these commercials might be tested by showing Version A in one market and Version B in a similar, but different, market. Actual sales of the product in both markets might then be monitored to determine which commercial was more successful in persuading viewers to buy the product. As can be seen, the results of the field experiment have more relevance to reality, but the degree of control involved is markedly less than is the case in the lab experiment.

Some field studies possess the advantage of being nonreactive. **Reactivity** is the influence that a subject's awareness of being measured or observed has on his or her behavior. Lab subjects are almost always aware of being measured. Although this is also the case in some field experiments, many can be conducted without subjects' knowledge of their participation.

Field experiments are useful for studying complex social processes and situations. In their study of the effects of the arrival of television in an English community, Himmelweit, Oppenheim, and Vince (1958) recognized the advantages of the field experiment for examining such a complicated topic. Since television has an impact on several lifestyle variables, the researchers employed a wide range of analysis techniques, including diaries, personal interviews, direct observation, questionnaires, and teachers' ratings of students, to document this impact. A topic area as broad as this does not easily lend itself to laboratory research.

Field experiments can be inexpensive. Most studies require no special equipment or facilities. However, expenses increase rapidly with the size and scope of the study (Babbie 1979).

Finally, the field experiment may be the only research option to use. For example, suppose a researcher is interested in examining patterns of communication at a television station before and after a change in management—a problem difficult if not impossible to simulate in a laboratory. His or her only practical option is to conduct the study in the field, that is, at the station.

Disadvantages of Field Experiments

The disadvantages of the field experiment are mostly practical ones. However, some research is impossible to conduct because of ethical considerations. The vexing question of the effects of television violence on young viewers provides a good example of this problem. Probably the most informative study that could be performed in this area would be a field experiment in which one group of children is required to watch violent television programs while another, similar group views only nonviolent programs. These two groups could be carefully observed over a number of years to check for any significant difference in the number of aggressive acts committed by the members of each group. Unfortunately, the ethics involved in controlling the television viewing behavior of children and possibly encouraging aggressive acts, are extremely questionable. Therefore, scientists have had to resort to laboratory and survey techniques to study this problem.

On a more practical level, field experiments often encounter external hindrances that cannot always be anticipated. For example, a researcher may spend weeks planning a study to manipulate the media usage of students in a summer camp, only to have the camp counselors or a group of parents decide that they do not want the children used as "guinea pigs" and scuttle the project. Also, it takes time for researchers to establish contacts, secure cooperation, and gain necessary permissions before beginning a field experiment. In many cases this phase of the process may take weeks or months to complete.

Finally, and perhaps most importantly, researchers cannot control all the intervening variables in a field experiment. The presence of these extraneous variables affects the precision of the experiment and the confidence the researchers have in its outcome.

Types of Field Experiments

There are two basic categories of field experiments: (1) those in which the researcher manipulates the independent variable(s); and (2) those in which independent variable manipulation occurs naturally as a result of other circumstances. To illustrate the first type, suppose that a researcher is interested in investigating the effects of not being able to read a newspaper. A possible approach would be to select two comparable samples and not allow one to read any newspapers for a period of time; the second sample (the control group) would continue to read the newspaper as usual. A comparison could then be made to determine if abstinence from newspapers has any effect in other areas of life, such as interpersonal communication. In this example, the researcher manipulates an independent variable (reading the newspaper).

The second type of field experiment involves passive manipulation of independent variables. Suppose a community with no cable television system is scheduled to be wired for cable in the near future. In an attempt to gauge the effects of cable on television viewing and other media use, a researcher might begin studying a large sample of television set owners in the community long before the cable service is available. A few months after it is introduced, the researcher could return to the original sample, sort out those households that subscribed to cable and those that did not, and proceed from there to determine the effects of the cable service. In this case, there is no control over the independent variable (cable service); the researcher is merely taking advantage of existing conditions.

Examples of Field Experiments

Tan (1977) was interested in what viewers would do if they were no longer permitted to watch television. Accordingly, he recruited a sample of 51 adults and paid each of them $4.00 every day not to watch television for an entire week. Before depriving these subjects of television, Tan requested that they watch television normally for a one-week period and keep a detailed diary of their activities. At the start of the experimental week, Tan's assistants visited the subjects' homes and taped up the electrical plugs on their television sets to lessen temptation. Again, the subjects were requested to record their activities for the week. In order to maintain some control over the experiment, the assistants visited the subjects' homes periodically during the week to ensure that television was not being viewed.

One week later, the diaries completed during the week of deprivation were collected and the data compared to the week of normal television viewing. Tan discovered that when deprived of television, subjects turned more to radio and newspapers for entertainment and information. They also tended to engage in more social activities with their friends and family.

This study illustrates some of the strengths and weaknesses of field experiments. In the first place, they are probably the only viable technique available to investigate this particular topic. A survey (see Chapter 7) does not permit the researcher to control whether the subjects watch television, and it would be impossible in the United States to select a representative sample composed of people who do not own a television set. And bringing people into the laboratory for a whole week of television deprivation makes even less sense.

On the other hand, the ability of the field experimenter to control independent variables is not conclusively demonstrated here: Tan had no way to assure that his sample actually avoided television for the entire week. The subjects could have watched at friends' homes or at local bars, or even at home by taking the tape off their plugs. Moreover, Tan mentioned that several individuals who fell into the initial sample refused to go without television for only $4.00 per day. As a result, the non-probability sample did not accurately reflect the general make-up of the community.

Two rather ambitious field experiments were conducted by Milgram and Shotland (1973) with the cooperation of the Columbia Broadcasting System television network. The researchers arranged to have three different versions of the popular television series *Medical Center* constructed. One version depicted antisocial behavior that was punished (a jail sentence); another portrayed antisocial behavior that went unpunished; and a third contained prosocial (favorable) behavior. The antisocial behavior consisted of scenes of a distraught young man smashing a plastic charity collection box and pocketing the money.

In the first experiment, the researchers recruited subjects in two ways: ads were placed in New York City newspapers that promised a free transistor radio to anyone willing to view a one-hour television show; and business reply cards containing the same message were passed out to pedestrians near several subway stops. Subjects were asked to report to a special television theater to view the program; upon arrival, each was randomly assigned to one of four groups, and each group was shown a different program (the three described above plus a different non-violent show used as a control). After viewing the program (with no commercial interruptions) and completing a short questionnaire about it, the subjects were instructed to go to an office in a downtown building to receive their free radio.

The downtown office was part of the experiment. The office was monitored by hidden cameras and contained a plastic charity collection box with around $5.00 in it—similar to the one seen in the *Medical Center* episode—and a note informing the subjects that no more transistor radios were available. Their behavior upon reading the note was to be the dependent variable: how many would emulate the antisocial act seen in the program and take the money from the charity box? Milgram and Shotland found no differences in antisocial behavior among the viewers of each group; no one broke into the charity box.

The second study tried to gauge the immediate effects of televised antisocial acts upon viewers as they watched. Subjects were recruited from the streets of New York City's Times Square area and ushered into a room with a color television set and a plastic charity collection box containing $4.45. A hidden camera monitored the subjects' behavior, even though they were told that they would not be observed. Once again, no differences emerged between the groups.

These two studies also demonstrate several positive and negative aspects of field experiments. In the first place, Milgram and Shotland had to secure the cooperation of CBS to conduct this expensive study. Second, volunteer subjects were used (see Chapter 4), and it is reasonable to assume that the sample was unrepresentative of the general population. Third, in the first experiment, the researchers did not control for the amount of time that passed between viewing the program and arriving at the testing center. Some participants arrived 24 hours after watching *Medical Center*, while others came several days later. Clearly, the subjects' experiences during this interval may have influenced their responses. Finally, Milgram and Shotland reported that the second experiment had to be terminated early because some of the subjects started resorting to behavior that the researchers could not control.

On the plus side, the first experiment clearly shows the potential of the field experiment to simulate natural conditions and to provide a nonreactive setting. Upon leaving the theater after seeing the program, subjects had no reason to believe that they were still participating in another phase of the research. Consequently, their behavior at the gift center was probably genuine and not a reaction to the experimental situation.

The Milgram and Shotland studies also raise the important question of ethics in field experiments. Subjects were observed without their knowledge and were apparently never told about the real purpose of the study, nor even that they were involved in a research study. Does the use of a hidden camera constitute an invasion of privacy? Does the experimental situation constitute entrapment? How about the subjects who stole the money from the "charity" box? Have they committed a crime? Field experiments can sometimes pose difficult ethical considerations, and these points must be dealt with *before* the experiment is conducted, not afterwards, when harm may already have been inflicted on the subjects (see Chapter 17).

Field Observations

Field observation has been used only rarely in mass media research. Only about 2–3% of the articles published in journalism and broadcasting journals have employed the technique (Lowry 1979) (although its use is a bit more common in advertising studies). Nonetheless, the technique holds promise for mass media research and should not be excluded offhand as a potential approach.

Field observation is useful for collecting data as well as for generating hypotheses and theories; it is more concerned with description and explanation than it is with measurement and quantification. There are two extremes of field observation, total observation and total participation. **Total observation** occurs when the observer assumes no role in the phenomenon being observed (subjects may or may not be aware that they are being studied). At the other extreme, **total participation** occurs when the observer becomes a full-fledged participant in the situation (subjects are usually unaware that a study is being conducted, although they may be told so).

To illustrate the distinction between the two approaches, consider the example of a researcher who wishes to observe and analyze the dynamics of comedy-writing for television. The researcher has two basic choices: to study television comedy writers in a passive role by merely watching them in action, or to participate directly in the process by joining the writing team. In the latter case, the researcher would probably want to be considered a fellow writer by the other members of the team (such arrangements might be made with the chief writer, who would be the only person to know the true identity of the researcher).

A combination of the two approaches, known as participant observation, occurs when a researcher is identified as such but still interacts with the other participants in the social or working situation. In the previous example, for instance, the researcher could have been introduced as a mass media investigator but still participate in the creative writing process.

The choice of technique depends on the research problem. Total participation may affect subjects' behavior and also raises the ethical question of deception. On the other hand, the information gathered may be more valid if subjects are unaware of being scrutinized.

Advantages of Field Observations

Field observation is not an appropriate technique for every research question, owing to the lack of control and quantification, but it does possess several unique advantages. For one thing, many mass media problems and questions cannot be studied using any other methodology. Field observation often helps the researcher to define basic background information necessary to frame a hypothesis and to isolate independent and dependent variables. For example, a researcher interested in how creative decisions in advertising are made could observe several decision-making sessions to see what actually transpires. Field observations often make excellent pilot studies in that they identify important variables and provide useful preliminary information.

A field observation is not always used as a preliminary step to other approaches, however. There are many cases where it alone is the only appropriate approach, especially in situations were quantification is difficult. A study of the gatekeeping (information flow) process in a network television news department is an instance where a field observation is particularly suitable: quantification of gatekeeping is rather tenuous.

Field observation may also provide access to groups that would otherwise be difficult to observe or examine. For example, a questionnaire sent to a group of people who produce X-rated movies is not apt to have a high return rate. An observer, however, may be able to establish enough mutual trust with such a group to convince them to respond to rigorous questioning.

Field observation is usually inexpensive. In most cases, field observations require only writing materials or a small tape recorder. Expenses increase if the problem under study requires a large number of observers, extensive travel, or special equipment (such as video recording machines).

Perhaps the most noteworthy advantage of field observation is that the study takes place in the natural setting of the activity being observed and can thus provide data rich in detail and subtlety. Many mass media situations, such as a family watching television, are complex and are constantly subjected to intervening influences. Field observation, because of the opportunity for careful examination, allows observers to identify these otherwise unknown variables.

Disadvantages of Field Observations

On the negative side, field observation is a bad choice if the researcher is concerned with external validity. This difficulty is partly due to the potentially questionable representativeness of the observations made and partly to problems in sampling. Observing the television viewing behavior of a group of children at a day-care center can provide valuable insights into the social setting of television viewing, but it probably has little correlation to what preschoolers do in other places and under different circumstances.

Another problem is that field observation relies heavily upon a researcher's perceptions and judgments as well as on preconceived notions about the material under study. Consequently, this experimenter bias may unavoidably favor specific preconceptions of results, while observations to the contrary are ignored or distorted. This is the primary reason why one observer is rarely used in a field observation study. Observations need to be *cross-validated* by a second or third observer.

Field Observation Techniques

There are at least four different stages in a typical field observation study: (1) gaining access; (2) sampling; (3) data collection; and (4) data analysis.

Access. The first step in gaining access to a setting and/or group of people is to define the research goal. Is there a need to observe homes with school-age children in them, or to watch newspaper reporters as they perform their daily routines? Once the goal has been clearly articulated, the next step is to establish contact. Observation of a formal group (such as a film production crew) often requires permission from management and perhaps union officials. School systems and other bureaucracies usually have a special unit to handle requests from researchers and to assist them in obtaining necessary permissions.

Gaining permission to conduct field observation research requires persistence and public relations skills. Researchers must determine how much to disclose about the nature of the research. In most cases it is not necessary to provide a complete explanation of the hypothesis and procedures, unless there may be objections to sensitive areas. Researchers interested in observing which family member actually controls the television set might explain that they are studying patterns of family communication. Once the contact is made, it is then necessary to establish a rapport with the subject(s). Bogdan and Taylor (1975) suggested the following techniques for building rapport: establish common interests with the participants; start relationships slowly; if appropriate, participate in common events and activities; and do not disrupt participants' normal routines.

Sampling. Sampling in field observation is more ambiguous than in most other research approaches. In the first place, there is the problem of how many individuals or groups to observe. If the focus of the study is communication in the newsroom, how many newsrooms should be observed? If the topic is family viewing of television, how many families should be included? Unfortunately, there are no guidelines to help answer these questions. The research problem and the goals of the study are often used as indicators for sample size: if the results are intended for generalization to a population, then one subject or group is probably inadequate.

Another problem is deciding what behavior episodes or segments to sample. The observer cannot be everywhere and see everything, so what is observed becomes a de facto sample of what is not observed. If an observer views one staff meeting in the newsroom, this meeting represents other, unobserved meetings; one conversation at the coffee machine is a sample of all such conversations. In many cases researchers cannot adhere closely to the principles of probability sampling, but they should keep in mind the general notion of representativeness.

Most field observations use purposive sampling: observers draw upon their knowledge of the subject(s) under study and sample only from those behaviors or events that are relevant. In many cases, previous experience and study of the activity in question will suggest what needs to be examined. In a study of newsroom decision-making, for example, researchers would want to observe staff meetings, since they are obviously an important part of the process. However, restricting the sampling to observations of staff meetings would be a mistake; many decisions are made at the water fountain, over lunch, and in the hallways. Experienced observers tend not to isolate a specific situation but rather to consider even the most insignificant situation for potential analysis. For most field observation, researchers need to spend some time simply getting the "feel" of the situation and absorbing the pertinent aspects of the environment before beginning a detailed analysis.

Data collection. The traditional tools of data collection—the notebook and pen—have given way to radically new equipment in many cases, due to recent advances in electronics. For example, Bechtel, Achelpohl, and Akers (1972) installed television cameras in a small sample of households in order to document the families' television-viewing behavior. Two cameras were automatically activated when the television set was turned on and videotaped the scene in front of the set. However, while a camera is able to record more information than an observer with a notebook, Bechtel reported that problems in finding consenting families, maintenance of the equipment, and low light levels made the project difficult.

Note-taking in the total participant situation requires special attention. Continually scribbling away on a notepad is certain to draw attention and suspicion to the researcher and might expose his or her real purpose in a particular setting. In a situation of this type, it is advisable to make mental notes and transcribe them at the first available opportunity. If the researcher is initially identified as such, the problem of note-taking is somewhat alleviated. Nonetheless, it is not recommended that the observer spend all of his or her time furiously taking notes. Subjects are already aware of being observed, and conspicuous note-taking could make them more uneasy. Brief notes taken during natural breaks in a situation attract a minimum of attention and can be expanded at a later time.

The field notes constitute the basic corpus of data in any field study. In them, the observers record not only what happened and what was said, but also personal impressions, feelings, and interpretations of what was observed. A general procedure is to separate personal opinions from the descriptive narrative by enclosing them in brackets.

How much should be recorded? It is always better to record too much information than too little. An apparently irrelevant observation made during the first viewing session might become significant during the course of the project. If the material is sensitive, or if the researcher wishes to keep secret the fact that research is taking place, then the notes may be written in abbreviated form or in a special code.

Data analysis. Quantitative data analysis procedures have only partial relevance to field observations due to the qualitative nature of field data. In field observation, data analysis consists primarily of filing and content analysis.

Constructing a filing system is an important step in observation. The purpose of the filing system is to take raw field data and arrange them in an orderly format to enable systematic retrieval at a later time (the precise filing categories are determined by the data). Using the hypothetical study of decision-making in the newsroom, filing categories might include the headings "Relationships," "Interaction—Horizontal," "Interaction—Vertical," and "Disputes." An observation may be placed in more than one category. It is a good idea to make multiple copies of all notes. In addition, a periodic filing of notes throughout the observation period will save time and confusion later on.

A rough content analysis is performed to search for consistent patterns once all the notes have been ascribed to their proper files: Perhaps most decisions in the newsroom are made in informal settings such as hallways rather than in formal settings such as a conference room. Perhaps most decisions are made with little superior-subordinate consultation. At the same time, deviations from the general norm should be investigated: Perhaps all reporters except one are typically asked their opinions on the newsworthiness of certain events. Why the exception?

The overall goal of data analysis in field observation is to arrive at a general understanding of the phenomenon under study. In this regard, the observer has the advantage of flexibility. In laboratory and other research approaches, investigators must at some point commit themselves to a particular design or questionnaire. If it subsequently turns out that a crucial variable was left out, there is little that can be done. In field observation, the researcher can analyze data during the course of the study and change the research design accordingly.

Some examples of field observation studies in mass media research include Geiber's (1956) classic study of gatekeeping in the newsroom and Epstein's (1974) description of network news operations. Ravage (1977) was able to gain access to all meetings, production work, and postproduction operations of a network company in his study of television production. Similarly, Pekurny (1980) focused on the production of the National Broadcasting Company's program *Saturday Night Live.* He was given access to the blocking, rehearsal, and live broadcast phases of the program. Pekurny also functioned as a participant observer when he took part in discussions with writers about how to structure a joke and about the suitability of some material for broadcast.

Focus Groups

The **focus group,** or group interviewing, is a research strategy typically used by advertisers and market researchers to help them understand consumer attitudes and behavior. However, the method is becoming very popular with broadcasters who are interested in finding out more about their audience.

The focus group technique involves interviewing two or more people simultaneously, with a moderator or facilitator leading the respondents in a relatively free discussion about the topic under consideration. The identifying characteristic of the focus group is the free-form discussion among the respondents; simultaneous gathering of data by questionnaire, for example, does not constitute a focus group. Focus groups can be used independently to collect purely qualitative data or in conjunction with other data collection methods.

Advantages of Focus Groups

The primary advantage of focus groups is the ability to collect preliminary information about a topic or phenomenon. Rarely are focus groups used as the only research approach because sample size is usually too small. Instead, focus groups are used as pilot studies to detect ideas that will be investigated further using another research method, such as a survey. Focus groups can be conducted in a short time span and are *usually* inexpensive—the price can range from a few hundred dollars to several thousand dollars, depending on the size of the group, the difficulties in sample selection, and the company that conducts the group.

Focus groups are also flexible (Babbie 1979). In conventional surveys, interviewers work from a rigid schedule of questions and are instructed to follow their directions exactly (see Chapter 7). A moderator in a focus group, however, works from a list of broad questions as well as more refined *probe* questions, so that he or she is always able to follow up on important points raised by participants in the group.

Lastly, focus group responses are often more complete and less inhibited than those of individuals. One respondent's remarks tend to stimulate another to pursue a line of thinking that might not have been brought out in an individual situation. With competent leadership, the discussion can have a snowballing effect, as one respondent comments on the views of another. Bold respondents tend to encourage more timid respondents to speak when they might ordinarily have remained silent. If one respondent opens up a topic or makes an admission, the barrier is broken and others may join in. Thus, more information is gathered than would be the case in a more structured situation. A skilled moderator can also detect the opinions and attitudes of those who are not particularly articulate by noting their facial expressions and nonverbal behavior while others are presenting their points of view.

Disadvantages of Focus Groups

Focus group research is not free from complications. Some sessions are dominated by a self-appointed group leader who monopolizes the conversation and attempts to inflict his or her opinion on the other participants; resentment is a natural outgrowth, and the atmosphere in the interview situation may become unpleasant. Moreover, interview findings may not be generalizable to the larger population, since sample sizes are relatively small.

Gathering hard data is inappropriate for a focus group. If quantification is important, it is wise to supplement the focus group with other research data where more specific questions can be addressed to a more representative sample. Many people unfamiliar with focus group research incorrectly assume that the method will answer questions of "how many" or "how much." Focus group research is intended to gather *qualitative* data—to answer questions such as "why" and "how." Many times broadcasters who hire a person or company to conduct a focus group are disgruntled with the results because they expected exact numbers and percentages. Focus groups do not provide such information; the procedure is a complement to other research techniques and must be interpreted in that light.

Another significant problem is the dependence of focus groups upon the skills of the moderator. This person must know when to probe for further information, when to stop respondents from discussing irrelevant topics, and how to get all respondents involved in the conversation (Fletcher and Wimmer 1981). Other drawbacks to the technique include the fact that the small samples involved are volunteers; the potential that recording equipment or other physical characteristics of the location will inhibit respondents; and the possibility that the data generated may not be useful if respondents are allowed to stray from the primary topic of the study.

The Methodology of Focus Groups

There are seven basic steps involved in focus group research (Fletcher and Wimmer 1981):

1. *Define the problem.* This step is similar in all types of scientific research: a well-defined problem is established, either on the basis of some previous investigation or out of curiosity to learn about something.

2. *Select a sample.* Because focus groups are small (usually between 6 and 12 participants), it is necessary for researchers to define a narrow audience to be involved in the study. The type of sample depends on the purpose of the focus group: the sample might consist of housewives who use a particular type of laundry detergent, men aged 18–34 who listen to a certain kind of music, or teenagers who purchase more than ten record albums per year.

3. *Determine the number of groups necessary.* In order to eliminate part of the problem of selecting a representative group, most researchers conduct two or more focus groups on the same problem. Results can then be compared to determine if any similarities exist, or one group can be used as a basis for comparison. A focus group study using only one group is rare, since there is no way to know if the results are group-specific or characteristic of a wider audience.

4. *Prepare the study mechanics.* This includes arranging for recruitment of focus group participants (by telephone or some other method) and for procurement of the facilities where the study will be conducted and any necessary recording equipment that will be used. In addition, a moderator must be selected and trained to conduct the sessions.

5. *Prepare the focus group materials.* Each aspect of the study must be planned in detail: the recordings or other materials that subjects will hear or see must be prepared, questionnaires developed, and an interviewer or moderator guide constructed. The general procedure in focus groups is to begin the session with some type of shared experience, so that the individuals have some common base from which to start the discussion. This might take the form of listening to a prerecorded tape or examining a new product. The interviewer's

guide should include the broad questions as well as the probe questions that will guide the group.

6. *Conduct the session.* Focus groups may be conducted in a variety of settings, from professional conference rooms equipped with two-way mirrors to allow researchers to observe the session, to motel rooms rented for the purpose of the study.

7. *Analyze the data and prepare a research report.* The analysis of focus group interviews depends on the needs of the study and the amount of time and money available. At one extreme, it may be enough for the moderator to write a short summary of what was said and provide an interpretation of subjects' responses. For a more elaborate content analysis, interviews can be taped and transcribed so that the researcher can carefully scan the comments and develop a relevant category system, coding each comment into the appropriate category. For example, a researcher who notices that most respondents focus on the price of a new product could establish a content category labeled "Price" and codes all statements in the transcript referring to price. These statements can then be arranged under the general heading. The same technique is followed for other content categories. When the coding is completed, the researcher makes summary statements about the number, tone, and consistency of the comments that fall into each category. Needless to say, this approach requires some expenditure of time and money on the researcher's part.

For more information on focus groups, see Fletcher and Wimmer (1981); Calder (1977); Elliot (1980); Cox, Higginbotham, and Burton (1976); and Reynolds and Johnson (1978).

Panel Studies

The panel design, discussed briefly in Chapter 5, is a method whereby the same sample of respondents is measured at different points in time. A panel study is basically a field research approach, although it can also make use of mail questionnaires, telephone or personal interviews, and lab techniques. Television networks, advertising agencies, and marketing research firms use panel studies to track changes in consumer behavior, such as trends, new ideas and fads, and buying habits. Panel studies on the effectiveness of political commercials, for example, interview all the members of the panel at periodic intervals during the campaign to determine if and when each subject makes a voting decision.

Depending on the purpose of a study, researchers can use either a *continuous panel,* consisting of members who report specific attitudes or behavior patterns on a regular basis, or an *interval panel,* whose members agree to complete a certain number of measurement instruments (usually questionnaires) only when the information is needed.

Advantages of Panel Studies

Panel studies can reveal shifting attitudes and patterns of behavior that might go unnoticed with other research approaches. These can include shifts in almost any type of variable: voting, products purchased, media usage, and so forth. Panel studies produce data suitable for sophisticated statistical analysis and enable researchers to predict cause-and-effect relationships. **Path analysis** and **cross-lagged studies** (see Chapter 12) are two techniques that have been used with panel data (Lefkowitz et al. 1972; Neale 1972).

Finally, panel studies help solve the problems normally encountered when defining a theory on the basis of a one-shot case study. Since the research progresses over a period of time, the researcher can allow for the influences of competing stimuli on the subject.

Disadvantages of Panel Studies

Panel members are often difficult to recruit because of unwillingness to fill out questionnaires or submit to interviews several times. The number of initial refusals in a panel study fluctuates depending upon the amount of time required, the prestige of the organization directing the study, and the presence or absence of some type of compensation. One analysis of the refusal rates in twelve marketing panel studies found a range of 15–80%, with a median of about 40% (Carman 1974).

Once the sample is secured, the problem of mortality emerges. Some panel members will drop out for one reason or another. Since the strength of panel studies lies in interviewing the same people at different times, this advantage diminishes as the sample size decreases. Another serious problem is that respondents often become sensitized to measurement instruments after repeated interviewing, thus making the sample atypical (see Chapter 4). For example, panelists who know in advance that they will be interviewed about public television watching might alter their viewing patterns to include more Public Broadcasting System programs (or fewer). Respondent error is always a problem in situations that depend on self-administered measurement instruments. If panelists are asked to keep a diary for a certain period of time, it is possible that they will forget to fill it out until immediately before it is due. Lastly, of course, panel studies require a long time investment and can be quite expensive.

Perhaps the most famous example of the panel technique in mass media research is the collection of national television audience data by the A. C. Nielsen company. Nielsen's sample consists of about 1,170 households spread across the United States. These homes are wired with a mechanical device called an audimeter that records when the television set is turned on and to which channel. Strictly speaking, the Nielsen sample is a panel of television sets rather than people. In order to collect information on viewers, Nielsen supplements this technique with diaries. Twenty percent of the Nielsen panel is supposedly replaced each year.

Other panels are maintained by such commercial research organizations as Market Facts, Inc., National Family Opinion, Inc., and the Home Testing Institute. In the non-marketing area, an ambitious panel study was carried out under the supervision of the National Broadcasting Company (Milavsky and Pekowsky 1973). In this study, a sample of 650 boys was measured 6 different times over a 3½ year period on a number of variables related to television viewing and aggressive behavior. The results indicated that both aggressiveness and viewing of violent television programs showed many fluctuations over the course of the study; the panel design highlighted the ebb and flow of these behavior patterns.

Summary

Although other research techniques are used more often than the ones discussed in this chapter, field research should not be overlooked by mass media investigators. As mentioned in the chapter, the limits of field research are basically limits of creativity, not of methods.

This chapter discussed four types of field research: field experiments, field observation, focus groups, and panel studies. Each of these methods was discussed with regard to its advantages and disadvantages. The decision to employ one or more of these techniques depends on the nature of the research problem and the types of data the researcher wishes to generate.

Questions and Problems for Further Investigation

1. Develop a research topic that would be appropriate for a study by:
   a. Field experiment
   b. Field observation
2. Suggest three specific research topics that would be best studied by the technique of total participation. Would any ethical problems be involved?
3. Select a research topic that is suitable for study using the focus group method, then assemble six or eight of your classmates and/or friends and conduct a sample interview. Select an appropriate method for analyzing the data.
4. Try to develop an unobtrusive measurement technique that would examine each of the following concepts:
   a. Newspaper reading
   b. Aggressive tendencies
   c. Brand loyalty
   d. Television viewing

References and Suggested Readings

Babbie, E. R. 1979. *The Practice of Social Research.* Belmont: Wadsworth.

Bechtel, R., Achelpohl, C., and Akers, R. 1972. "Correlates between Observed Behavior and Questionnaire Responses on Television Viewing." *Television and Social Behavior,* Vol IV, ed. by E. Rubinstein, G. Comstock, and J. Murray. Washington, D.C.: U.S. Government Printing Office.

Bickman, L., and Hency, T. 1972. *Beyond the Laboratory: Field Research in Social Psychology.* New York: McGraw-Hill.

Bogdan, R., and Taylor, S. 1975. *Introduction to Qualitative Research Methods.* New York: Wiley.

Calder, B. J. 1977. "Focus Groups and the Nature of Qualitative Marketing Research." *Journal of Marketing Research* 14:353–64.

Carman, J. M. 1974. "Consumer Panels." *Handbook of Marketing Research,* ed. by R. Ferber. New York: McGraw-Hill.

Cox, K. D., Higginbotham, J. B., and Burton, J. 1976. "Applications of Focus Group Interviewing in Marketing." *Journal of Marketing* 40:77–80.

Elliot, S. C. 1980. "Focus Group Research: A Workbook for Broadcasters." Washington, D.C.: National Association of Broadcasters.

Epstein, E. J. 1974. *News from Nowhere.* New York: Vintage.

Fletcher, A., and Bowers, T. 1979. *Fundamentals of Advertising Research.* Columbus: Grid.

Fletcher, J. E., and Wimmer, R. D. 1981. *Focus: Group Interviews in Radio Research.* Washington, D.C.: National Association of Broadcasters.

Geiber, W. 1956. "Across the Desk: A Study of 16 Telegraph Editors." *Journalism Quarterly* 33:423–32.

Haskins, J. B. 1968. *How to Evaluate Mass Communications.* Chicago: Advertising Research Foundation.

Himmelweit, H., Oppenheim, A. N., and Vince, P. 1958. *Television and the Child.* London: Oxford University Press.

Lefkowitz, M., Eron, L., Walder, L., and Huesmann, L. 1972. *Television and Adolescent Aggressiveness,* ed. by E. Rubinstein, G. Comstock, and J. Murray. Washington, D.C.: U.S. Government Printing Office.

Lowry, D. 1979. "An Evaluation of Empirical Studies Reported in Seven Journals in the '70s." *Journalism Quarterly* 56:262–68.

Milavsky, J. R., and Pekowsky, B. 1973. "Exposure to TV Violence and Aggressive Behavior in Boys, Examined as Process: A Status Report of a Longitudinal Study." Unpublished paper, Department of Social Research, National Broadcasting Company.

Milgram, S., and Shotland, R. 1973. *Television and Antisocial Behavior: Field Experiments.* New York: Academic Press.

Neale, J. 1972. "Comment on Television Violence and Child Aggression: A Follow-up Study." *Television and Social Behavior,* Vol. III, ed. by G.

Comstock and E. Rubinstein. Washington, D.C.: U.S. Government Printing Office.

Pekurny, R. 1980. "The Production Process and Environment of NBC's *Saturday Night Live.*" *Journal of Broadcasting* 24:91–100.

Ravage, J. W. 1977. "Not in the Quality Business: A Case Study of Contemporary Production." *Journal of Broadcasting* 21:47–60.

Redding, C. W. 1970. "Research Settings: Field Studies." In P. Emmert, *Methods of Research in Communication*, ed. by P. Emmert and W. D. Brooks. New York: Houghton Mifflin.

Reid, L. N., Soley, L. C., and Wimmer, R. D. 1981. "Replication in Advertising Research: 1977, 1978, 1979." *Journal of Advertising* 10:3–13.

Reynolds, F. D., and Johnson, D. K. 1978. "Validity of Focus Group Findings." *Journal of Advertising Research* 18:21–24.

Tan, A. 1977. "Why TV is Missed: A Functional Analysis." *Journal of Broadcasting* 21:371–80.

Tukey, J. 1962. "The Future of Data Analysis." *Annals of Mathematical Statistics* 33:1–67.

Tull, D., and Hawkins, D. 1980. *Marketing Research*. New York: MacMillan.

Westley, B. H. 1981. "The Controlled Experiment." *Research Methods in Mass Communication*, ed. by G. H. Stempel and B. H. Westley. Englewood Cliffs: Prentice-Hall.

Wimmer, R. D., and Reid, L. N. 1982. "Willingness of Communication Researchers to Respond to Replication Requests." *Journalism Quarterly* (in press).

CHAPTER 7

Survey Research

Surveys are now used in all areas of life. Businesses, consumer groups, politicians, and advertisers use them as part of their everyday decision-making processes. Some firms, such as Gallup, Harris, and Frank N. Magid Associates, conduct surveys on a full-time basis.

The importance of survey research to the public at large is confirmed by the frequent reporting of survey results in the popular media. This is especially the case during campaign periods, when the public continually hears or reads about various popularity polls conducted to ascertain candidates' positions with the electorate.

The increased use of surveys has created changes in the way they are conducted and reported. More attention is now given to sample selection, questionnaire design, and error rates. This means that surveys require careful planning and execution; mass media studies using survey research must take into account a wide variety of decisions and problems. This chapter acquaints the beginning researcher with the basic steps of survey methodology.

Descriptive and Analytical Surveys

At least two major types of surveys are used by researchers: descriptive and analytical. A **descriptive survey** attempts to picture or document current conditions or attitudes, that is, to *describe* what exists at the moment. For example, the Department of Labor regularly conducts surveys on the status of labor unemployment in the United States. Professional pollsters survey the electorate to learn its opinions of candidates or the issues of a campaign. Broadcast stations and networks continually survey the viewing audience to determine programming tastes, changing values, and lifestyle variations that might affect programming. In descriptive surveys of this type, researchers are interested in discovering the current situation in a given area.

Analytical surveys attempt to describe and explain *why* certain situations exist. This approach usually involves an examination of two or more variables in order to test research hypotheses. The results allow researchers to examine the interrelationships among variables and draw explanatory inferences. For example, television station owners occasionally survey their market to determine how lifestyles affect viewing habits, or to determine if viewers' lifestyles can be used to predict the potential success of syndicated programming. On a much broader scale, television networks conduct yearly surveys to determine how the public's tastes and desires are changing and how these attitudes relate to the perception each viewer has of the three commercial networks.

Analytical and descriptive surveys are important to mass media researchers in both the academic and the private sector. Syndicated survey researchers such as Simmons and A. C. Nielsen provide media management with current audience data. In addition, custom research, designed by companies like Frank N. Magid Associates and McHugh-Hoffman, allows broadcasters to investigate unique problems. Academicians, on the other hand, generally conduct theoretical rather than applied research. For example, Chaffee and Choe (1980) investigated those voters in presidential elections who choose a candidate during, rather than prior to or near the end of, the campaign. Poindexter (1979) examined non-readers of daily newspapers and found that they consist of two distinct groups. And Wakshlag and Greenberg (1979) studied the relationship between television programming strategies and the popularity of children's programs.

Advantages and Disadvantages of Survey Research

Surveys have five well-defined advantages. First, surveys can be used to investigate problems in realistic settings. Newspaper reading, television viewing, and consumer behavior patterns can be examined where they happen, rather than in a laboratory or screening room under artificial conditions.

Second, the cost of surveys is reasonable considering the amount of information gathered. In addition, researchers can control expenses by selecting from three major types of surveys: mail, telephone, and personal interview.

A third advantage is that large amounts of data can be collected with relative ease from a variety of people. The survey technique allows the researcher to examine a large number of variables (demographic and lifestyle information, attitudes, motives, intentions, and so on) and to use multivariate statistics (see Chapter 12) to analyze the data. Also, geographic boundaries do not limit most surveys.

Finally, data helpful to survey research already exist. Data archives, government documents, census materials, rating books, and voter registration lists can be used as *primary* sources (main sources of data) or *secondary* sources (supportive data) of information. It is possible to conduct an entire survey study without ever developing a questionnaire or contacting a single respondent (Wimmer 1976).

Survey research is not a perfect research methodology. The technique also possesses several disadvantages. The first and most important is that independent variables cannot be manipulated as in laboratory experiments. Without control of independent variable variation, the researcher cannot be certain whether the relations between independent and dependent variables are causal or noncausal. That is, a survey may establish that A and B are related, but it is impossible to determine solely from the survey results that A causes B. Causality is difficult to establish because many intervening and extraneous variables are involved. Time series studies may help correct this problem, but not always.

A second disadvantage is that the wording of questions and the placement of items within questionnaires can have biasing effects on survey results. Great care must be taken to assure that the data collected as well as the measurement process are reliable and valid. This problem is discussed later in the chapter.

Another potential disadvantage of survey research is its dependence on sampling techniques. If the sample is largely unrepresentative, then the results will have little relevance to other situations, even though the sample size may be quite large (see Chapter 4).

Survey research has both advantages and disadvantages, but they do produce reliable and useful information. They are especially useful for collecting information on audiences and readership; however, they are not particularly suited for testing hypotheses.

Constructing Questions

Constructing good survey questions involves two basic considerations: (1) the questions must clearly and unambiguously communicate the desired information to the respondent; and (2) the questions should be worded to allow accurate transmission of respondents' answers to researchers.

Questionnaire design depends on the choice of a data collection technique. Questions written for a **mail survey** must be easy to read and understand, since respondents are unable to ask questions. **Telephone surveys** cannot use questions with long lists of response options; the respondent may forget the first few responses by the time the last ones are read. In a **personal interview** the interviewer must tread lightly with sensitive and personal questions which his or her physical presence might make the respondent less willing to answer. (These procedures are discussed in greater detail later in this chapter.)

In designing a questionnaire, the basic purpose of the research must always be considered. A complex research topic such as media use during a political campaign requires more detailed questions than does a survey to determine a favorite radio station or magazine. Nonetheless, there are several general guidelines to follow in questionnaire design regarding wording of questions and question order and length.

Types of Questions

Surveys consist of two basic types of questions, open-ended and close-ended. An **open-ended question** is one in which respondents are asked to provide their own answers to a question. For example:

What do you like most about your local newspaper?

____

____

What type of television program do you prefer?

____

____

What three problems are the most important in your community?

____

____

____

Open-ended questions allow respondents freedom in answering questions and the chance to provide in-depth responses. Furthermore, they give researchers the opportunity to ask: "Why did you give that particular answer?" or "Could you explain your answer in more detail?" This flexibility to follow up on, or probe, certain questions enables the interviewers to gather information about the respondents' feelings and the motives behind their answers.

Also, open-ended questions allow for answers that researchers did not originally foresee in the construction of the questionnaire and that may suggest possible relationships with other answers or variables. For example, in response to the question, "What types of programs would you like to hear on radio?" the manager of a local radio station might expect answers such as "news" or "weather" or "sports." However, a subject may give an unexpected response such as "obituaries" (Fletcher and Wimmer 1981). This would force the manager to reconsider his perceptions of some of the local radio listeners.

Finally, open-ended questions are particularly useful in pilot versions of major studies. Researchers may not know what types of responses to expect from subjects, so open-ended questions can be used to allow subjects to answer in any way they wish. From the list of responses provided by the subjects, the researchers can then select the most-often mentioned items and include them in multiple choice or forced choice questions. Using open-ended questions on a pilot study generally saves time and resources since all possible responses are more likely to be included on the final measurement instrument; there would be no reason to reconduct the analysis for failure to include an adequate number of responses or response items.

The major disadvantage to open-ended questions is the amount of time it takes to collect and analyze the data. Since subjects' responses are always varied, it is necessary to analyze their content (see Chapter 8) before they can be tabulated. Content analysis allows common responses to be grouped into categories, making the question similar to a forced choice question (see below).

In the case of **close-ended questions,** respondents are asked to select an answer from a list provided by the researcher. These questions are popular both because they provide greater uniformity of response and because the answers are easily quantifiable for computer analysis. The chief disadvantage of close-ended questions is that researchers may fail to include some important responses. Respondents do not have the opportunity to answer questions according to their own beliefs or feelings and sometimes comment that the "correct" responses were not available. One common technique is to include a response of "other" followed by a blank space. This provides subjects an opportunity to include an answer not mentioned in a question's response. Unfortunately, many respondents simply check "other" without writing in their preferred response. A pilot study can help determine which responses to include.

General Observations

Before examining specific types of questions appropriate in survey research, some general do's and don'ts about writing questions are in order.

1. *Make questions clear.* This should go without saying, but many researchers become so closely associated with a problem that they can no longer put themselves in the respondents' position. What might be perfectly clear to researchers might not be nearly as clear to persons answering the question. For example, "What do you think of our company's rebate program?" might seem a perfectly sensible question to a researcher, but to respondents it might mean, "Is the monetary amount of the rebate too small?" "Is the rebate given on the wrong items?" "Does it take too long for the rebate to be paid?" "Have the details of the program been poorly explained?" and so on. Questionnaire items must be precise so that respondents know exactly what is being asked.

Making questions clear also requires avoiding difficult or specialized words, acronyms, and stilted language. In general, the level of vocabulary commonly found in newspapers or popular magazines is adequate for a survey. Questions should be phrased in everyday speech, and social science jargon and technical words should be eliminated. For example, "Would you consider buying a VHS or Beta system if it retailed for less than $500?" might be better phrased: "Would you buy a machine that could record and play back television programs if it cost less than $500?" The item, "Should the city council approve the construction of an interactive CATV system?" assumes that respondents know what "interactive CATV systems" are. A preferable option is: "An interactive cable television system is one in which viewers can send messages back to the cable company as well as receive normal television. Do you think the city council should approve such a system for this community?"

The clarity of a questionnaire item can be affected by double or hidden meanings in the words that are not apparent to investigators. For example, the question, "How many television shows do you think are a little too violent—most, some, few or none?" contains such a problem. Some respondents will answer "none," even though they feel that TV shows are extremely violent, on the basis of the question wording "some." These subjects reason that all shows are more than "a little too violent"; therefore, the most appropriate answer to the question is "none."

2. *Keep questions short.* In order to be precise and unambiguous, researchers sometimes write long and complicated items. However, respondents are often in a hurry to complete a questionnaire, and it is unlikely that they will take the time to study the precise

intent of items. Short, concise items that will not be misunderstood are best.

3. *Include complete instructions.* Questionnaire instructions vary depending on the type of survey conducted. Mail surveys usually require the most specific instructions, since respondents are not able to ask questions about the survey. Respondents and interviewers should understand whether the correct response should be circled, checked, placed in a specific order, or skipped. **Filter questions,** which are used to eliminate unwanted subjects, often require respondents or interviewers to skip one or more questions. This information should be clearly specified. For example:

Do you listen to FM radio?

_____ YES	_____ NO
(If yes, go to Q 2)	(If no, skip to Q 5)

A survey using this question might be designed only to question subjects who listen to FM radio. The filter question immediately assesses whether the subject falls into this group. If the subject responds "NO," then the interviewer skips a certain number of related questions, or may terminate the interview at this time since the subject is unwanted.

Some questionnaires require respondents to rank a list of items. In this case, the instructions must clearly describe which response represents the highest value:

Please rate the following magazines in order of importance. Place a "1" next to the magazine you prefer most, a "2" next to the magazine in second place, and so on up to "5."

_____ PLAYBOY

_____ BETTER HOMES AND GARDENS

_____ POPULAR SCIENCE

_____ READER'S DIGEST

_____ SCIENTIFIC AMERICAN

All questions should be tested in a pilot study to determine if directions for answering questions are clear.

4. *Remember the purposes of the research.* It is important to include only those items in a questionnaire that directly relate to what is being studied. For example, if the occupational level of the respondents is not relevant to the hypothesis, the questionnaire should not ask about it. Beginning researchers often tend to add questions merely for the sake of developing a longer questionnaire. Keep in mind that parsimony in questionnaires is a paramount consideration.

5. *Do not ask double-barreled questions.* A **double-barreled question** is one that actually asks two or more questions. Whenever the word *and* appears in a question, the sentence structure should be

examined to see if more than one question is being asked. For example, "This product is mild on hands and gets out stubborn stains. Do you agree ____ or disagree ____?" It is possible that the product might get out stubborn stains and at the same time be highly irritating to the skin; a respondent could agree with the second part of the question while disagreeing with the first part. This question should be divided into two items.

6. *Avoid biased words or terms.* Consider the following item: "In your free time, would you rather read a book or just watch television?" The word *just* in this example injects a pro-book bias into the question because it implies that there is something less than desirable about watching television. In like manner, "Where did you hear the news about the president's new program?" is mildly biased against newspapers; the word *hear* suggests that "radio," "television," or "other people" is a more appropriate answer. Questionnaire items that start off with, "Do you agree or disagree with so-and-so's proposal to . . ." almost always bias a survey. If the name "Adolf Hitler" is inserted for "so-and-so," the item becomes overwhelmingly negative. By inserting "the President," a potential for both positive and negative bias is created. Any time a specific person or source is mentioned in a question, the possibility of bias arises.

7. *Avoid leading questions.* A **leading question** is one that suggests a certain response (either literally or by implication) or contains a hidden premise. For example, "Like most Americans, do you read a newspaper everyday?" suggests that the respondent should answer in the affirmative or run the risk of being unlike most Americans. The question, "Do you still use marijuana?" contains a hidden premise. This type of question is usually referred to as a *double bind:* regardless of how the respondent answers, an affirmative response to the hidden premise is implied (that is, he or she has used marijuana at some point).

8. *Do not use questions which ask for unusually difficult or detailed information.* The question, "In the last 30 days, how many hours of television have you viewed with your family?" is unrealistic. Few respondents could answer such a question. A more realistic approach would be to ask, "How many hours did you spend watching television with your family yesterday?" If the researcher is interested in a 30-day period, he or she should ask respondents to keep a log or diary of family viewing habits.

9. *Avoid potentially embarrassing questions unless absolutely necessary.* Most surveys need to collect data of a confidential or personal nature, but an overly personal question may cause embarrassment and inhibit respondents from answering honestly. Two common areas with high potential for embarrassment are age and income. Many individuals are reluctant to tell their exact ages to strangers doing a survey. Instead of asking directly how old a respondent is, it is better to allow him or her some degree of confidentiality by asking,

"Now, about your age, . . . are you in your 20s, 30s, 40s, 50s, 60s. . . ?" Most respondents are willing to state what decade they fall in, and this information is usually adequate for statistical purposes. Interviewers might also say, "I'm going to read several age categories to you. Please stop me when I reach the category you're in."

Income may be handled in a similar manner. A straightforward, "What is your annual income?" often prompts the reply, "None of your business." It is more prudent to ask, "Which of these categories includes your total annual income?"

_____ More than $30,000	_____ $4,000–$7,999
_____ $15,000–$29,999	_____ $2,000–$3,999
_____ $8,000–$14,999	_____ Under $2,000

These categories are broad enough to allow respondents some privacy but narrow enough to allow for statistical analysis. Moreover, the bottom category, "Under $2,000," was made artificially low so that individuals would not have to be embarrassed by checking the $2,000–$3,999 slot. The income classifications depend on the purpose of the questionnaire and the geographic and demographic distribution of the subjects. The $30,000 upper level in the example may be much too low in several parts of the country.

Other potentially sensitive areas include people's sex lives, drug usage, religion, business practices, and trustworthiness. In all of these areas, care should be taken to ensure respondents of confidentiality and even anonymity, when possible.

The simplest type of close-ended question is one that provides a *dichotomous response,* usually "agree-disagree" or "yes-no." For example,

Television stations should editorialize.

_____ AGREE

_____ DISAGREE

_____ NO OPINION

While such questions provide little sensitivity to different degrees of conviction, they are the easiest to tabulate of all question forms. Whether they provide enough sensitivity is a question the researcher must seriously consider.

The *multiple choice* question allows respondents to choose an answer from several options. For example:

In general, television commercials tell the truth . . .

_____ ALL OF THE TIME

_____ MOST OF THE TIME

_____ SOME OF THE TIME

_____ RARELY

_____ NEVER

Multiple choice questions should include all possible responses. A question that excludes response usually creates problems. For example:

What is your favorite television network?

_____ ABC

_____ CBS

_____ NBC

Subjects who favor PBS (although not a *network* in the strictest sense of the word) cannot answer the question as presented.

Additionally, multiple choice responses must be **mutually exclusive:** there should be only one response option per question for each respondent. For instance:

How many years have you been working in newspapers?

_____ LESS THAN ONE YEAR

_____ ONE TO FIVE YEARS

_____ FIVE TO TEN YEARS

Which blank should a person with exactly five years of experience check?

One way to correct this problem is to reword the responses, such as:

How many years have you been working in newspapers?

_____ LESS THAN ONE YEAR

_____ ONE YEAR TO FIVE YEARS

_____ SIX TO TEN YEARS

Rating scales are also widely used in mass media research (see Chapter 2). They can be arranged horizontally or vertically:

There are too many commercials on TV.

_____ STRONGLY AGREE (translated as a "5" for analysis)

_____ AGREE (translated as a "4")

_____ NEUTRAL (translated as a "3")

_____ DISAGREE (translated as a "2")

_____ STRONGLY DISAGREE (translated as a "1")

What is your opinion of TV news?

FAIR _____ _____ _____ _____ _____ UNFAIR

(5) (4) (3) (2) (1)

Semantic differential scales (see Chapter 2) are another form of rating scale and are frequently used to rate persons, concepts, or objects. These scales use bipolar adjectives with seven scale points:

How do you perceive the term *public television*?

GOOD	_____	_____	_____	_____	_____	_____	_____	BAD
HAPPY	_____	_____	_____	_____	_____	_____	_____	SAD
UNINTERESTING	_____	_____	_____	_____	_____	_____	_____	INTERESTING
DULL	_____	_____	_____	_____	_____	_____	_____	EXCITING

In many instances researchers are interested in the relative perception of several concepts or items. In such cases the *rank ordering* technique is appropriate:

Here are several common occupations. Please rank them in terms of their prestige. Put a "1" next to the profession that has the most prestige, a "2" next to the one with the second most, and so on.

_____ POLICEMAN _____ TV REPORTER

_____ BANKER _____ TEACHER

_____ LAWYER _____ DENTIST

_____ POLITICIAN _____ NEWSPAPER WRITER

Ranking is not recommended with more than a dozen objects because the process can become tedious and the discriminations exceedingly fine. Furthermore, ranking data imposes limitations on the statistical analyses that can be performed.

The **checklist** question is often used in pilot studies to refine questions for the final project. For example:

What things do you look for in a new television set? (Check as many as apply.)

_____ AUTOMATIC FINE TUNING _____ CONSOLE MODEL

_____ REMOTE CONTROL _____ PORTABLE

_____ LARGE SCREEN _____ STEREO SOUND

The most frequently checked answers are later used to develop a multiple choice question; the unchecked responses are dropped.

The *forced choice* question is frequently used in media studies designed to gather information about lifestyles and are always listed in pairs. Forced choice questionnaires are usually very long—sometimes hundreds of questions—and repeat questions (in different form) on the same topic. The answers for each topic are analyzed for patterns. A respondent then receives a score for his or her interest in that topic. A typical forced choice questionnaire might look like this:

Select one statement from each of the following pairs of statements:

_____ I enjoy attending parties with my friends.

_____ I enjoy staying at home alone.

_____ Gun control is necessary to stop crime.

_____ Gun control can only increase crime.

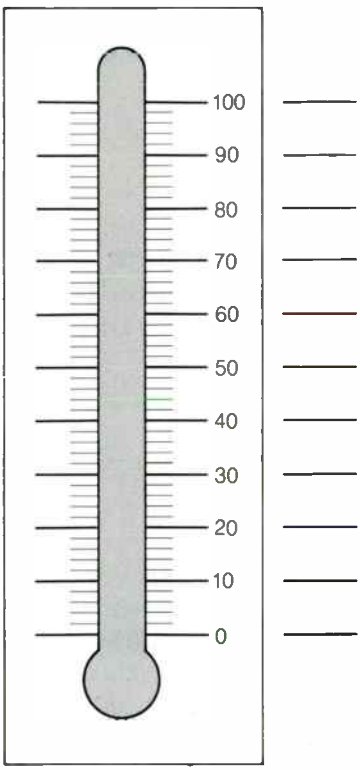
_____ If I see an injured animal, I always try to help it.

_____ If I see an injured animal, I figure that nature will take care of it.

Respondents generally complain that neither of the responses to a forced choice question is satisfactory, but they have to select one or the other. Through a series of questions on the same topic (violence, lifestyles, career goals), a pattern of behavior or attitude generally develops.

Fill-in-the-blank questions are used infrequently by survey researchers. However, some types of studies are particularly suited for the fill-in-the-blank questions. In advertising copy testing, for example, they are often employed to test subjects' recall of a commercial. After seeing, hearing, or reading a commercial, subjects receive a script of the commercial in which a number of words have been randomly omitted (often every fifth or seventh word). Subjects are required to fill in the missing words to complete the commercial. Fill-in-the-blank questions can also be used in information tests. For example, "The Senators from your state are _________ and _________." Or, "The headline story on the front page was about _______."

Tables, graphs, and figures are also used in survey research. Some ingenious questioning devices have been developed in order to

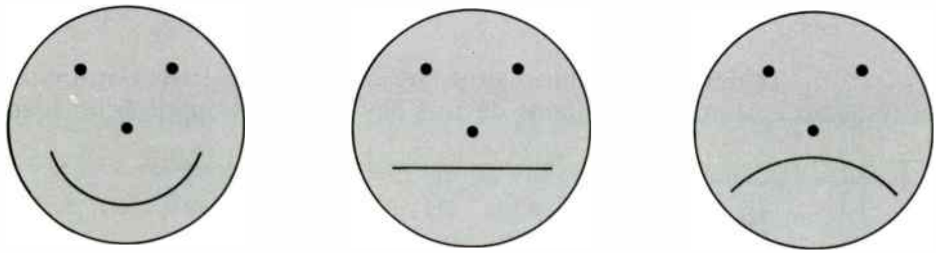


help respondents more accurately describe how they think and feel. For example, the University of Michigan Survey Research Center developed the "feeling thermometer," on which subjects can rate an idea or object. The thermometer, which is patterned after a normal mercury thermometer, is an easy way for subjects to rate their degree of like or dislike in terms of "hot" or "cold." For example:

> How would you rate the coverage your local newspaper provided on the recent school board campaign? (Place an "X" near the number on the thermometer which most accurately reflects your feelings; a "100" indicates strong approval, and a "0" reflects strong disapproval.)

Some questionnaires designed for children use other methods to collect information. It is well known that young children have difficulty in assigning numbers to values; one logical alternative is to use pictures. For example:

> How do you feel about Saturday morning cartoons on television?



1 2 3

Questionnaire Design

The approach used in asking questions as well as the physical appearance of a questionnaire can affect the response rate. The time and effort involved in developing a good questionnaire always pay off with a larger sample size. The following section offers some suggestions concerning these two points.

Introduction

One way to increase response rate in any type of survey is to prepare a persuasive introduction to the survey. It is important to convince respondents early on that the survey is a legitimate research project. It is also necessary to convince them that the survey is valuable and will help solve problems, whether theoretical or applied. In a telephone survey it is wise to tell respondents how much time is needed to complete the interview: "This will only take about three minutes of your time. . . ." A mail survey introduction might similarly stress the brevity of the task: ". . . a one-page questionnaire is enclosed."

Regardless of the survey approach used, a well-constructed introduction will usually generate higher response rates than a simple, "Please answer the following questions. . . ."

Instructions

All instructions necessary to complete the questionnaire should be clearly stated for both respondents and interviewers. Procedural instructions for respondents are often highlighted using a different typeface or some other typographical device, perhaps arrows or lines. The following is an example from a mail survey:

> Do you have a favorite radio station that you listen to most of the time?
>
> _____ YES➡ _____ NO
>
> If yes, can you remember the names of any of the disc jockeys or newscasters who work for that station? *WRITE THE NAMES BELOW.*
>
> ______________________________
>
> ______________________________
>
> ______________________________
>
> ______________________________

Procedural instructions for interviewers, on the other hand, are usually distinguished from the rest of the questionnaire by being typed in capital letters and often enclosed in boxes. For example:

> We'd like to start by asking you some things about television. First, what are your favorite TV shows?
>
> | TAKE DOWN ALL NAMES OF TV SHOWS. GET SPECIFIC NAMES, NOT CATEGORIES OF SHOWS. PROBE WITH "ARE THERE ANY MORE?" TO GET AT LEAST THREE SHOWS. |
> |---|
>
> 1. ____________ 3. ____________
>
> 2. ____________ 4. ____________

When interviewers are used, as is the case with telephone and person-to-person surveys, it is imperative that they be supplied with an easy-to-follow questionnaire. All instructions should be clear and simple. A confusing questionnaire impairs the effectiveness of the interviewer, lowers the number of respondents who complete the test, and, in the long run, increases costs.

Question Order

All surveys tend to flow better when the initial questions are simple and easy to answer. Researchers often include one or two "warm-up" questions not directly related to the study so that respondents become accustomed to the task of answering. Preliminary questions can also serve as motivational steps to create subject interest in the questionnaire. Demographic data, personal questions, and other sensitive items should be placed toward the end of the questionnaire to allow the interviewer to establish a rapport with the respondents. Subjects may still refuse to answer these questions, but at least the main body of data will already have been collected.

The questionnaire should be organized in a logical sequence, proceeding from the general to the specific. Questions on similar topics should be grouped together, and the transitions between different question sections should be clear and logical.

Poor question order may bias a respondent's answers. For example, suppose several questions are asked about the presence of violence in society; then the respondent is asked to rank the major problems facing the country today from the following list:

_____ WAR	_____ HIGH PRICES
_____ COMMUNISM	_____ CORRUPT GOVERNMENT
_____ VIOLENCE ON TV	_____ POLLUTION

It is possible that violence on television might receive a higher ranking than it would if the ranking question had been asked before the series of questions on violence. Or, to take another example, suppose a public relations researcher is attempting to discover the public's attitudes toward a large oil company. If the questionnaire began with attitudinal questions concerning oil spills and inflated profits and then asked respondents to rate certain oil companies, it is likely that the ratings of all the companies would be lower, due to general impressions created by the earlier questions.

There is no easy solution for the problem of question "contamination." Obviously, some questions have to be asked first, some second, and so on. Perhaps the best thing researchers can do is to be sensitive to the problem and test for it in a pilot study. If they think that question order A, B, C will have biasing effects, they should administer another version using the order C, B, A so that an objective positioning of the questions can be determined.

Layout

The physical design of the questionnaire is another important factor in survey research. A badly typed, poorly reproduced questionnaire is not apt to attract many responses in a mail survey. Nor does a cramped questionnaire with 40 questions to a page help to instill respondents with a positive attitude. Response categories should be adequately spaced and presented in a non-confusing manner. For example, the following format might lead to problems:

There are too many commercials on television.

DO YOU STRONGLY AGREE ____ AGREE ____ HAVE NO OPINION ____ DISAGREE ____ STRONGLY DISAGREE ____?

A more effective and less confusing method is to provide a vertical ordering of the response choices:

There are too many commercials on television.

____ STRONGLY AGREE

____ AGREE

____ NO OPINION

____ DISAGREE

____ STRONGLY DISAGREE

Some researchers recommend avoiding blanks altogether because respondents and interviewers tend to make large check marks or X's that cover more than one blank, making interpretation difficult. If blanks are perceived as a problem, one can provide boxes to check or numbers to circle. In any case, the response form should be consistent throughout the questionnaire. Format changes generally create confusion for both respondents and interviewers.

Finally, each question must have enough space for answers. This is especially true for open-ended questions. Nothing is more discouraging to respondents and interviewers than to be confronted with a situation like:

Why do you go to the movies? ____________

Who are your favorite movie stars? ____________

What is your favorite television show? ____________

If a research budget does not allow for a long questionnaire (created by adding extra white space for answers), then subjects should be invited to add further comments on the back of the survey or on another sheet of paper.

Pilot Studies

Without a doubt, the best way to discover if any research instrument is adequately designed is by pretesting it. That is, conduct a pilot study to check for problems. Small samples can be invaluable for determining if the study approach is correct and for refining questions. Areas where misunderstanding or confusion occur can be easily corrected without wasting time or money.

Pilot study results provide researchers with information to rework the questionnaire: to delete or rewrite questions, to change open-ended questions to close-ended questions (or vice versa), and to verify that all response options have been provided. A second run-through is recommended if several problems are uncovered in the first test.

The pilot study should also include a run-through for interviewers. A "dress rehearsal" of the full-scale project insures that the interviewers fully understand what is being researched.

Gathering Survey Data

After developing a questionnaire and conducting one or more pilot studies, the next step is to gather data from a sample of respondents. There are three basic methods for doing this: (1) the mail survey; (2) the telephone survey; and (3) the personal interview survey. Researchers can also use variations and combinations of these three methods. Each procedure has definite advantages and disadvantages to be considered before a choice is made. The remainder of this chapter highlights the good and bad points of each method.

Mail Surveys

Mail surveys involve mailing self-administrable questionnaires to a sample of individuals. Prestamped reply envelopes are enclosed to allow respondents to mail the completed questionnaire back to the researcher. Mail surveys are popular because they can secure a great deal of data with a minimum expenditure of time and money. At the outset, however, researchers should be aware that respondents are busy people with many demands on their time. Consequently, many people do not share the researcher's enthusiasm for questionnaires and often simply throw them away.

The general stages of a mail survey are discussed below. Even though the steps are listed in numerical sequence, many of these tasks are often accomplished in a different order or even simultaneously.

1. *Select a sample.* Sampling is generally done from a prepared frame (see Chapter 4) that contains the names and addresses of potential respondents. The most common sampling frame used is the **mailing list**, a compilation of names and addresses in narrowly defined groupings that commercial firms sometimes prepare (see boxed material below).

2. *Construct the questionnaire.* As discussed earlier, mail survey questionnaires must be concise and specific, since no interviewer is present to alleviate misunderstandings, answer questions, or give directions.

Obtaining Specialized Lists

More than 50 firms specialize in providing mailing lists to advertisers and marketing research organizations. A few of the more prominent companies and some of the types of lists they provide are listed below:

R. R. Bowker Company
1180 Avenue of the Americas
New York, NY 10036
(bookstores; colleges; libraries; librarians; audiovisual specialists)

Customized Mailing Lists
158–23 Grand Central Parkway
Jamaica Estates, NY 11432
(business and professional people; executives)

Dunhill International List Company
2430 W. Oakland Park Boulevard
Fort Lauderdale, FL 33311
(engineers; lawyers; physicians; small business owners)

Market Development Company
41 Kimler Drive
Hazlewood, MO 63043
(families with children; newlyweds; high school students)

Zeller and Letica, Inc.
15 E. 26th Street
New York, NY 10010
(clergy; retailers; ethnic names)

3. *Write a cover letter.* A brief note explaining the purpose and importance of the questionnaire usually increases response rates.

4. *Assemble the package.* The questionnaires, cover letters, and return envelopes are stuffed into mailing envelopes. Researchers sometimes choose to use bulk mail with first-class return envelopes. An alternate method is to send questionnaires first class and use business reply envelopes for responses. This method allows researchers to pay only for those questionnaires actually returned. Postal options always depend on the research budget.

5. *Mail the surveys.*

6. *Closely monitor the return rates.*

7. *Send follow-up mailings.* The first follow-up should be sent two weeks after the initial mailing, and a second (if necessary) two weeks after the first. The follow-up letters can be sent to the entire sample or only those subjects who failed to answer.

8. *Tabulate and analyze the data.*

Advantages. Mail surveys cover a wide geographic area for a rather reasonable cost. They are often the only way to gather information from people who live in hard-to-reach areas of the country (or in other countries). Mail surveys also allow for selective sampling through the use of specialized mailing lists. In addition to those mentioned above, lists are available that include only those people with an annual income over $50,000, or those consumers who have bought a car within the past year, or those who subscribe to a particular magazine, or those who live in a specific zip code area. If researchers need to collect information from a highly specialized audience, then the mail technique can be quite attractive.

Another advantage of the mail survey is that it provides anonymity, so that subjects are more likely to provide candid answers to sensitive questions. Questionnaires can be completed at home or in the office and thus provide the subject with a certain sense of privacy. People can answer questions at their own pace and have an opportunity to look up certain facts or check past information. Mail surveys also eliminate interviewer bias, since there is no personal contact involved.

Probably the biggest advantage of this method, however, is its relative low cost. Mail surveys do not require a large staff of trained workers. The only costs are for printing, mailing lists, envelopes, and postage. If the cost per completed questionnaire were to be computed, it is likely that mail surveys would prove the most inexpensive of all the survey methods.

Disadvantages. To begin with, mail questionnaires must be self-explanatory. There is no interviewer present to answer questions or to clear up misunderstandings. Mail surveys are also the slowest form of data collection. Returns may start to trickle in around a week or so after the initial mailing and continue to be received for several weeks thereafter. In fact, it may take months before some responses are returned. Many researchers simply set a cut-off date, after which returns are not included in the analysis.

Another problem with mail surveys is that researchers never know exactly who answered the questions. A survey sent to corporate executives, for example, may actually be answered by assistants. Furthermore, replies are often received only from those people who are interested in the survey, and this injects bias into the results. Most researchers agree, however, that the biggest disadvantage of the mail survey is the typically low return rate that goes with it. The "average" survey (depending upon the area and type of survey) will achieve a response rate of 10–25%. This low return casts doubt on the validity of the findings.

There are, however, a number of procedures for improving return rates that have been investigated by survey researchers. While there are no hard and fast guarantees, the techniques mentioned below have been shown to increase response rates in some mail surveys.

1. *Keep questions to a minimum.* One study (Sewell and Shaw 1968) found that while only 28% of the sample returned a three-page

questionnaire, 50% returned a double postcard containing only a single question. It would appear that up to a certain point, questionnaire length and response rate are negatively correlated.

2. *Use follow-ups.* The most common follow-up procedure is to send a letter (or postcard) to each respondent about two weeks after the initial mailing. In this letter, all those who have returned the questionnaire are thanked, while those who have not returned it are urged to do so. Occasionally, a second copy of the questionnaire is included with the follow-up letter. Using follow-up letters can produce an additional 10–20% return, and at least one study that used multiple follow-up requests reported a 50% increase. Of course, follow-up letters add to the total cost of the survey.

3. *Use inducements.* Money is the most common form of inducement used in survey research. The sum involved is not large and it is not designed to pay individuals for their answers, but is rather an acknowledgement that someone has gone out of his or her way to answer the questions. To illustrate, one study utilized an inducement of 25¢; the response rate among those who received the money was 52%, compared with 19% among those who received no reward (Erdos 1974).

4. *Include a cover letter.* An introductory letter stressing the importance of the survey can increase returns. While not as potent a factor as the other items mentioned above, a well-constructed cover letter might add an additional 10% or more to the response rate.

Other factors that have been shown to increase response rates include: (1) having an important person or prestigious group sponsor the survey; (2) using stamped envelopes rather than business reply envelopes; (3) asking for objective rather than subjective information; and (4) addressing cover letters to individuals by name, as opposed to "Dear Sir or Madam" (Miller 1977).

All in all, researchers who are willing to spend time, energy, and money in a mail survey can usually ensure an above-average return rate by utilizing one or more of these techniques.

Telephone Surveys

Telephone surveys and personal interviews both employ trained members of a research team to ask questions verbally and record the responses. The respondents generally do not get a chance to see the actual questionnaire. Since both telephone and personal interviewing techniques have certain similarities, much of what follows has relevance to personal interviews as well.

Telephone surveys seem to fill a middle ground between mail surveys and personal interviews. They offer more control and higher response rates than most mail surveys but are limited as to the types of questions that can be used. They are generally more expensive than mail surveys but less expensive than face-to-face interviews. Because of these factors, telephone surveys seem to represent a compromise between the other two techniques, and this may account for their growing popularity in mass media research.

Interviewers are extremely important to both telephone and personal surveys. An interviewer ideally should function as a neutral medium through which the respondents' answers are communicated to the researcher. The interviewer's presence and manner of speaking should not influence respondents' answers in any way. Adequate training and instruction can minimize bias that the interviewer might inject into the data. For example, if he or she shows disdain or shock over an answer, it is unlikely that the respondent will continue to answer questions in a totally honest manner. Showing agreement with certain responses might prompt similar answers to other questions. Skipping questions, carelessly asking questions, and being impatient with the respondent might also cause problems. As an aid to minimizing interviewer bias, the *National Association of Broadcasters* has published the following recommendations to interviewers.*

1. Read the questions exactly as worded. Ask them in the exact order listed. Skip questions only when the instructions on the questionnaire tell you to. There are no exceptions to this.
2. Never suggest an answer, try to explain a question, or imply what kind of reply is wanted. Don't prompt in any way.
3. If a question is not understood, say, "Let me read it again," and repeat it slowly and clearly. If it is still not understood, report a "no answer."
4. Report answers and comments exactly as given, writing fully. If an answer seems vague or incomplete, probe with neutral questions, such as, "Will you explain that?" or, "How do you mean that?" Sometimes just waiting a bit will tell the respondent you want more information.
5. Act interested, alert, and appreciative of the respondent's cooperation. But never comment on his or her replies. Never express approval, disapproval, or surprise. Even an "Oh" can cause a respondent to hesitate or refuse to answer further questions. Never talk up or down to a respondent.
6. Follow all instructions carefully, whether you agree with them or not.
7. Thank each respondent. Leave a good impression for the next interviewer.

A general procedure for conducting a telephone survey follows. Again, while the steps are presented in numerical order, it is possible to address many tasks simultaneously.

1. *Select a sample.* Telephone surveys require researchers to specify clearly the geographic area to be covered and to identify which persons will be interviewed in each household contacted. Many surveys are restricted to people over 18, heads of households, and so forth. The sampling procedure used depends on the purpose of the study (see Chapter 4).

*From *A Broadcast Research Primer*, 1976, pp. 37–38. Reprinted with permission.

2. *Construct the questionnaire.* Phone surveys require straightforward and uncomplicated response options. Ranking a long list of items is especially difficult over the telephone and should be avoided. In addition, the length of the survey should not exceed 10 minutes for nonprofessional interviewers. Longer interviews require professionals who are capable of keeping people on the telephone.

3. *Prepare an interviewer instruction manual.* This document should cover the basic mechanics of the survey (what numbers to call, when to call, how to record times, and so on). It should also specify which household member to interview (when to ask for the man or the lady of the house) and provide general guidelines on how to ask the questions and record the responses.

4. *Train the interviewers.* Interviewers need to practice going through the questionnaire to become familiar with all the items, response options, and instructions. It is best to train interviewers in a group using interview simulations and allowing each person to practice and ask questions. It is advisable to pretest interviewers as well as the questionnaire.

5. *Collect the data.* Data collection is most efficient when conducted from one central location (assuming enough telephone lines are available). Problems that develop are easier to remedy, and important questions raised by one interviewer that have general relevance can easily be communicated to the rest of the group. A central location also makes it easier for researchers to check (validate) the interviewers' work. The completion rate should also be monitored during this stage.

6. *Make necessary call-backs.* Additional calls (usually no more than two) should be made to those respondents whose lines were busy or who did not answer during the first session. Call-backs done on a different day or night tend to have a greater chance of success in reaching someone willing to be interviewed.

7. *Verify the results.* When all questionnaires have been completed, a small subsample of each interviewer's respondents should be called again to check that the information they provided was accurately recorded. Respondents should be told during the initial survey that they may receive an additional call at a later date. This tends to eliminate any confusion when subjects receive a second call. A typical procedure is to ask the subject's first name in the interview so that it can be used later. The interviewer should ask, "Was James called a few days ago and asked questions about television viewing?" The verification can begin from there, and need consist of only two or three of the original questions (preferably open-ended and sensitive questions, since interviewers are most apt to omit these).

8. *Tabulate the data.* Along with the normal data analysis, telephone researchers generally compute a response rate: how many com-

pleted interviews, how many refusals, how many no answers, and how many disconnects.

Advantages. The cost of telephone surveys tends to be reasonable. The sampling involves minimal expense, and there are no elaborate transportation costs. Call-backs are simple and economical. The Wide Area Telephone Service (WATS) enables researchers to conduct telephone surveys on a nationwide basis from any location.

Compared to mail surveys, telephone surveys can include more detailed questions, and, as stated earlier, interviewers can clarify misunderstandings that might arise during the administration of the questionnaire.

The non-response rate of a telephone survey is generally low, especially when multiple call-backs are employed. In addition, phone surveys are much faster than mail. A large staff of interviewers can collect the data from the needed sample in a relatively short time. In summary, phone surveys tend to be fast, easy, and relatively inexpensive.

Disadvantages. First of all, researchers must recognize that much of what is called survey "research" by telephone is not research at all, but an attempt to sell people something. Unfortunately, many companies disguise their sales pitch as a "survey," and this has made respondents suspicious and even prompted them to terminate an interview before it has gotten started. Additionally, visual questions are prohibited. A researcher cannot, for example, hold up a picture of a product and ask if the respondent remembers seeing it advertised. Lastly, a potentially severe problem is that not everyone in a community is listed in the telephone directory, the most often used sampling frame. Not everyone has a phone, and many people have unlisted phone numbers; also, some numbers are listed incorrectly, and others are too new to be listed. These problems would not be serious if the people with no phones or unlisted numbers were just like those listed in the phone book. Unfortunately, researchers generally have no way of checking for such similarities or differences, so it is possible that a sample obtained from a telephone directory may be significantly different from the population. (See Chapter 4 concerning random digit dialing.)

Personal Interviews

Personal interviews generally involve having the interviewer visit the respondent at his or her home or place of work. There are two basic types of interviews, structured and unstructured. In a **structured interview,** standardized questions are asked in a predetermined order; relatively little freedom is given to interviewers. In an **unstructured interview,** broad questions are asked, which allows interviewers freedom in determining what further questions to ask in order to obtain the required information. Structured interviews are easy to tabulate and analyze but do not achieve the depth or expanse of unstructured interviews. Conversely, the unstructured type elicits more detail but takes a great deal of time to score and analyze.

The steps involved in constructing a personal interview survey are similar to those for a telephone survey. In the list below, instances where the personal interview differs substantially from the telephone method are discussed.

1. *Select a sample.* Drawing a sample for a personal interview survey is more complicated than for a telephone survey. Usually a multistage procedure must be used. A common technique is to obtain detailed maps of the sample area, number each block, pick blocks at random, and then randomly select dwelling units within blocks (see Chapter 4).

2. *Construct the questionnaire.* Personal interviews are flexible: detailed questions are easy to ask, and the time taken to complete the survey can be greatly extended (many personal interviews take 30–60 minutes to complete). Researchers can also make use of visual exhibits, lists, and photographs to ask questions, and respondents can be asked to sort items into categories (the *Q-sort technique*) or to point to their answers on printed cards. Respondents can have privacy and anonymity by marking ballots which can then be slipped into envelopes and sealed.

3. *Prepare an interviewer instruction manual.* For the personal interview, this manual will need to be more detailed. Instructions should be included on what households to visit, who in the household to interview, what to do if a vacant dwelling falls into the sample, when to do the interviewing, how to dress and act during the interview period, how to record data, and so on.

4. *Train the interviewers.* Training is important because the questionnaires are longer and more detailed. Interviewers should receive instruction on gaining access and establishing a rapport with subjects; administrative details, such as when to conduct the interviews, how long each will take, and how much the interviewers will be paid; and follow-up questions. Several practice sessions are necessary to ensure that the goal of the project is being met, and that interviewers are following the established guidelines.

5. *Collect the data.* This part of the survey is the sole reason why many researchers select telephone or mail surveys. Interviewers' expenses for travel, food, and salary can be substantial (up to $100 per interview, or more with some types of private sector research; it is not uncommon for custom research companies to charge $1,000 per subject). Data collection also takes longer, possibly several months. Researchers need to consider these points before deciding to conduct personal interviews.

6. *Make necessary call-backs.* Each call-back requires an inter-

viewer to return to a household originally selected in the sample, and each call-back takes longer (and costs more) than in a phone survey.

7. *Verify the results.* As with the phone survey, a subsample of each interviewer's completed questionnaires is selected and assigned to another staff member. This person returns to the household in question (or calls on the telephone) and reviews a few of the more pertinent items to check on the original interviewer's completeness and veracity.

8. *Tabulate the data.* As with the phone survey, a completion rate should be computed, noting how many people refused to be interviewed, how many were not at home, and so forth.

Advantages. Many of the advantages of this technique have already been mentioned. It is the most flexible means of obtaining information, since the face-to-face situation lends itself easily to questioning in greater depth and detail. Further, some information can be observed by the interviewer during the interview without adding to the length of the questionnaire. For example, what magazines were displayed in the home? Was the television or radio turned on? Did the respondent have an outside antenna? Additionally, the interviewers can develop a rapport with the respondents and may be able to get replies to sensitive questions that would remain unanswered in a mail or phone survey.

The identity of the respondent is known or can be controlled in the personal interview survey. Whereas in a mail survey it is possible that all members of a family might confer on an answer, in a face-to-face interview, this can usually be avoided. Lastly, once an interview has begun, it is harder for respondents to terminate the interview before its completion. In a phone survey all the subject needs to do is to hang up the phone.

Disadvantages. The single largest drawback to the personal interview technique is cost. The transportation and labor costs involved generally make this the most expensive of all data collection methods. If respondents are spread out over a wide geographic area, it is easy to see that the travel, food, and lodging costs for interviewers could quickly mount up. Additionally, in recent years people have become more and more reticent to let strangers into their homes, even if these individuals have credentials from a reputable survey organization. It would appear that people are becoming more security conscious and more skeptical of individuals who represent themselves as "survey researchers." This fact may make it difficult to complete a large percentage of interviews.

Another major disadvantage is the problem of interviewer bias. The physical appearance, age, race, sex, dress, nonverbal behavior, and/or comments of the interviewer may prompt respondents to answer questions untruthfully. Moreover, the organization necessary for recruiting, training, and administering a field staff of interviewers is much greater than that required for other data collection procedures.

If large numbers of interviewers are needed, it is usually necessary to employ field supevisors to coordinate their work, which in turn will make the survey even more expensive. Finally, it should be noted that if personal interviewing is done during the day, most of the respondents will not be employed outside of the home. If it is desirable to interview those with jobs outside the home, it will be necessary to schedule interviews on the weekends or during the evening. This fact might make it more difficult to establish initial cooperation and drive costs even higher.

General Problems in Survey Research

Although surveys are valuable tools in mass communication research, there are problems present in any survey. Experience in survey research will show that:

1. Subjects or respondents are often unable to recall information about themselves or their activities. This inability may be caused by memory failure, confusion about the questions asked on the questionnaire, or some other intervening factor. Questions which seem glaringly simple to researchers may create severe problems for respondents. With this in mind, survey researchers must accept the fact that not all respondents will be able or willing to provide the required information.

2. **Prestige bias**—due to the respondent's feelings of inadequacy or lack of knowledge about a particular topic or area—is a phenomenon present in any type of survey. That is, respondents will often provide a "prestigious" answer rather than admit that they don't know something. Prestige bias is particularly apparent in surveys which probe respondents' television viewing habits: a respondent is apt to report heavy viewing of public television rather than confess an addiction to situation comedies or soap operas on commercial television. It is human nature not to want to feel "dumb" or "inferior," so answers may be given which have the air of sophistication or education.

3. Subjects may also purposefully deceive researchers by giving incorrect answers to a question. There is almost nothing which can be done about subjects who knowingly lie in their responses. A large enough sample may discount this type of response. At present, there are no acceptable and valid methods of determining whether a respondent's answers are truthful; the answers must be accepted as they are given.

4. Surveys are often complicated by the inability of respondents' to explain their true feelings and beliefs, not because they don't have any, but because they can't put them into words. The question, "Why do you like to watch soap operas?" may be particularly difficult for some respondents. They may watch them every day but respond

only with, "Because I like them." Probing respondents for further description may help, but not in every case.

Survey research can be an exciting process. Researchers must continually be aware of any obstacles which may hinder data collection and deal with these problems immediately. The United States is the most surveyed country in the world, and many citizens now refuse to take part in any survey study. Researchers must convince the subjects that the study will be helpful in making a particular decision or in solving of a particular problem.

Summary

Survey research is an important and useful method of data collection. The survey is also one of the most widely used methods of media research, primarily due to its flexibility. Surveys, however, involve a number of steps. Researchers must decide whether to use a descriptive or an analytical approach; define the purpose of the study; review the available literature in the area; select a survey approach, a questionnaire design, and a sample; analyze the data; and, finally, decide whether to publish or disseminate the results. These steps are not necessarily taken in that order, but all must be considered before a survey is conducted.

In order to ensure that all of the steps in the survey process are in harmony, researchers should conduct one or more pilot studies to detect any errors in the approach. Pilot studies save time, money, and frustration, since an error which could void the entire analysis might otherwise be overlooked.

Questionnaire design is also a major step in any survey. In this chapter, examples have been provided to show how a question or interviewing approach may elicit a specific response. The goal in questionnaire design is to avoid bias in answers. Question wording, length, style, and order may affect a respondent's answers. Extreme care must be taken when questions are developed to ensure that they are of a neutral nature.

Finally, researchers are charged with selecting a survey approach from among three basic types: mail, telephone, and personal interview. Each approach has advantages and disadvantages which must be weighed before a decision is made. The type of survey will depend on the purpose of the study, the amount of time available to the researcher, and the funds available for the study.

Questions and Problems for Further Investigation

1. Develop five questions or hypotheses which could be tested by survey research. What approaches could be used to collect information on these topics?

2. Non-response is a problem in all survey research. In addition, many people refuse to participate in surveys at all. Provide an example of a cover letter for a survey on television viewing habits. Include comments which might help increase the response rate.
3. Design questions to collect information on the following topics:
    a. Political party affiliation
    b. Attitudes toward television soap operas
    c. Attitudes toward newspaper editorials
    d. The frequency of television commercials
    e. Public television viewing habits
4. Locate one or more survey studies in journals related to mass media research. Answer the following questions in relation to the article(s):
    a. What was the purpose of the survey?
    b. How were the data collected?
    c. What type of information was produced by the survey?
    d. Did the data answer a particular research question or hypothesis?
    e. Were any problems evident with the survey and its approach?
5. Design a survey to collect data on a topic of your choice. Be sure to address the following points:
    a. What is the purpose of the survey? What is its goal?
    b. What research questions or hypotheses will be tested?
    c. Are any operational definitions required?
    d. Develop at least ten questions relevant to the problem.
    e. Describe the approach to be used to collect data.
    f. Design a cover letter or interview schedule for the study.
    g. Conduct a brief pilot study to test the questionnaire.

References and Suggested Readings

Babbie, E. R. 1973. *Survey Research Methods.* Belmont: Wadsworth.

Chaffee, S. H., and Choe, S. Y. 1980. "Time of Decision and Media Use During the Ford-Carter Campaign." *Public Opinion Quarterly* 44:53–70.

Erdos, P. L. 1974. "Data Collection Methods: Mail Surveys." *Handbook of Marketing Research,* ed. by R. Ferber. New York: McGraw-Hill Book Company.

Fletcher, J. E., and Wimmer, R. D. 1981. *Focus Group Interviews in Radio Research.* Washington, D.C.: National Association of Broadcasters.

Kerlinger, F. N. 1973. *Foundations of Behavioral Research.* 2nd ed. New York: Holt, Rinehart and Winston.

Miller, D. C. 1977. *Handbook of Research Design and Social Measurement.* New York: David McKay.

Oppenhiem, A. N. 1966. *Questionnaire Design and Attitude Measurement.* New York: Basic Books.

Poindexter, P. M. 1979. "Daily Newspaper Non-Readers: Why They Don't Read." *Journalism Quarterly* 56:764–70.

Rosenberg, M. 1968. *The Logic of Survey Analysis.* New York: Basic Books.

Sewell, W., and Shaw, M. 1968. "Increasing Returns in Mail Surveys." *American Sociological Review* 33:193.

Wakshlag, J. J., and Greenberg, B. S. 1979. "Programming Strategies and the Popularity of Television Programs for Children." *Journal of Communication* 6:58–68.

Walizer, M. H., and Wienir, P. L. 1978. *Research Methods and Analysis: Searching for Relationships.* New York: Harper & Row.

Weisberg, H. F., and Bowen, B. D. 1977. *An Introduction to Survey Research and Data Analysis.* San Francisco: Freeman.

Wimmer, R. D. 1976. "A Multivariate Analysis of the Uses and Effects of the Mass Media in the 1968 Presidential Campaign." Unpublished doctoral dissertation, Bowling Green University.

Winkler, R. L., and Hays, W. L. 1975. *Statistics: Probability, Inference and Decision.* 2nd ed. New York: Holt, Rinehart and Winston.

CHAPTER 8

Content Analysis

The chapters in Part Two up to this point have concentrated on more general approaches used in mass media investigation. This chapter moves to content analysis, a specific research approach used frequently in all areas of the media. The method is popular with mass media researchers because it provides an efficient way to investigate the content of the media such as the number and types of commercials or advertisements in broadcasting or in the print media. Beginning researchers will find content analysis a valuable tool in answering many mass media questions.

Modern content analysis can be traced back to World War II, when allied intelligence units painstakingly monitored the number and types of popular songs played on European radio stations. By comparing the music played on German stations with that on other stations in occupied Europe, the allies were able to measure with some degree of success the changes in troop concentration on the continent. In the Pacific theater, communication between Japan and various island bases was carefully tabulated; an increase in message volume usually meant that some new operation involving that particular base was planned. After the war, content analysis was used by researchers studying propaganda in newspapers and radio. In 1952, Bernard Berelson published *Content Analysis in Communication Research*, which signalled that the technique had gained recognition as a tool for communication scholars.

Since that time, the method has achieved wide popularity. In 1968, Tannenbaum and Greenberg reported that content analysis of newspapers was the largest single category of master's theses in mass communication. A 1975 publication (Comstock 1975) listed more than 225 content analyses of television programming. Recent concern over the portrayal of violence on television and the treatment of women and minority groups in both print and television advertising has further popularized the content analysis technique among mass media researchers.

Definition of Content Analysis

Many definitions of content analysis exist. Walizer and Wienir (1978) have defined it as any systematic procedure which is devised to examine the content of recorded information. Krippendorf (1980) defined it as a research technique for making replicable and valid references from data to their context. Kerlinger's (1973) definition is fairly typical: content analysis is a method of studying and analyzing communication in a systematic, objective, and quantitative manner for the purpose of measuring variables.

Kerlinger's definition involves three concepts that require elaboration. First, content analysis is *systematic*. This means that the choice of the content to be analyzed is made according to explicit and consistently applied rules: sample selection must follow proper procedures, and each item must have an equal chance of being included in the analysis. It also means that the evaluation process must be systematic: all content under consideration is to be treated in exactly the same manner. There must be uniformity in the coding and analysis procedures, as well as in the length of time coders are exposed to the material. Systematic evaluation simply means that one and only one set of guidelines for evaluation is used throughout the study. Alternating procedures in an analysis is a sure way to confound the results.

Second, content analysis is *objective*. That is, the personal idiosyncracies and biases of the investigator should not enter into the findings; the results of an analysis should be the same if replicated by another researcher. Operational definitions and rules for classification of variables should be explicit and comprehensive enough that other researchers who repeat the process will arrive at the same decisions.

Unless a clear set of criteria and procedures are established that fully explain the sampling and categorization methods, the researcher does not meet the requirement of objectivity, and the reliability of the results may be called into question. It should be noted, however, that perfect objectivity is seldom achieved in a content analysis. The specification of the unit of analysis and the precise make-up and definition of relevant categories are areas where individual researchers must exercise some subjective choice. (Reliability is discussed at length later in the chapter.)

Third, content analysis is *quantitative*. The goal of content analysis is the accurate representation of a body of messages. Quantification is important in fulfilling that objective, since it aids researchers in the quest for precision. Statements such as "Seventy percent of all prime time programs contain at least one act of violence" are more precise than, "Most shows are violent." Additionally, quantification allows researchers to summarize results and report them with greater parsimony. If measurements are to be made over intervals of time, comparisons of the numerical data from one time period to another can help to simplify and standardize the evaluation procedure. Lastly, quantification gives researchers additional statistical tools to use that can aid in interpretation and analysis.

Uses of Content Analysis

Over the past few years, the symbols and messages contained in the mass media have become increasingly popular research topics in both the academic and the private sector. The American Broadcasting Company conducts systematic comparative analyses of the three networks' evening newscasts to see how its news coverage relates to that of its competitors. The national chapter of the Parent-Teachers' Association has offered "do-it-yourself" training in rough forms of content analysis so that local members can monitor television violence levels in their viewing areas. Public relations firms use content analysis to monitor the subject matter of company publications, and some labor unions have recently begun performing content analyses of the mass media in order to examine their image.

Content analysis in the mass media often makes use of **medium variables,** those aspects of content that are unique to the medium under consideration. For example, in newspapers and magazines such variables might include typography, layout and make-up; in television, they might include shot duration, editing pace, shot selection, scene location, and camera angle. A discussion of medium variables as they relate to television news is contained in Adams and Schreibman (1978).

Although it is difficult to classify and categorize studies as varied and diverse as those using content analysis, they are generally employed for one of five purposes. A discussion of these aims will help illustrate some of the ways in which this technique can be applied.

Describing Communication Content

Several recent studies have catalogued the characteristics of a given body of communication content at one or more points in time. These studies exemplify content analysis used in the traditional, descriptive manner: to identify what exists. For example, Katzman (1972) described the problems, events, characters, and conversations involved in one week of soap operas. In like manner, Adams and Ferber (1977) described the positions and party affiliations of guests on Sunday television interview programs. Other analyses of this type describe trends in content over time, such as Greenberg et al.'s (1980) description of characters in three different seasons of prime time television.

Testing Hypotheses of Message Characteristics

A number of analyses attempt to relate certain characteristics of the source of a given body of message content to characteristics of the messages that are produced. As Holsti (1969) pointed out, this category of content analysis has been used in many studies that test hypotheses of form: "If the source has characteristic A, then messages containing elements x and y will be produced; if the source has characteristic B, then messages with elements w and z will be produced." Merritt and Gross (1978), for example, found that female editors of women's lifestyle pages were more likely than male editors to use stories about the women's movement. A study of local television newscasts revealed that the "eyewitness" format broadcast more news in the violent and human interest categories (Dominick, Wurtzel, and Lometti 1975).

Comparing Media Content to the "Real World"

Many content analyses may be described as "reality checks," in which the portrayal of a certain group, phenomenon, trait, or characteristic, is assessed against a standard taken from actuality. The congruence of the media presentation and the situation that actually exists is then discussed. Probably the earliest study of this type was Davis (1951) who found that crime coverage in Colorado newspapers bore no relationship to changes in state crime rates. DeFleur (1964) compared television's portrayal of the work world with actual job data taken from the U.S. census. More recently, the National Commission on the Causes and Prevention of Violence used content analysis data collected by Gerbner (1969) to compare the world of television violence with real life violence.

Assessing the Image of Particular Groups in Society

An ever growing number of content analyses have focused on exploring the media image of certain minority or otherwise notable groups. In many instances, these studies are conducted to assess changes in media policy toward these groups, to make inferences about the media's responsiveness to demands for "better" coverage, or to document social trends. For example, Hennessee and Nicholson (1972) performed an extensive analysis of the image presented of women by a New York television station as part of a license renewal challenge. Mertz (1970) analyzed the portrayal of the elderly on network television, and Greenberg and Kahn (1970) examined changes in the portrayal of blacks in *Playboy* cartoons.

Establishing a Starting Point for Studies of Media Effects

The use of content analysis as a starting point for subsequent studies is relatively new. The most well-known example is the **cultivation analysis**; whereby the dominant message and themes in media content are documented by systematic procedures, and a separate study of the audience is conducted to see if these messages are cultivating similar attitudes among heavy media users. Gerbner *et al.* (1979) discovered that heavy viewers of television (where crime and violence are popular elements) tend to be more fearful of the world around them. In other words, television content may foster attitudes more consistent with its messages than with reality. Other studies that have used a similar framework include DeFleur and DeFleur's (1967) study of the possible effects of occupational stereotypes in television programs and Dominick's (1973) study of the influence of television crime on the viewers' perception of actual crime.

Limitations of Content Analysis

It should be emphasized that content analysis alone cannot serve as a basis for making statements about the effects of content on an audience. A study of Saturday morning cartoon programs on television might reveal that 80% of these programs contain commercials for sugared cereal, but this fact alone does not allow researchers to claim that children who watch these programs will want to purchase sugared cereals. To make such an assertion, an additional study of the viewers would be necessary (as in cultivation analysis). By itself, content analysis cannot be used to make claims about media effects.

Also, the findings of a particular content analysis are limited to the framework of the categories and definitions used in that analysis. Different researchers may use varying definitions and category systems to measure a single concept. In mass media research, this problem is most evident in studies of televised violence. Some researchers rule out comic or slapstick violence in their studies, while others consider it an important dimension. Obviously, great care must be exercised in comparing the results of different content analysis studies. Researchers who use different tools of measurement will naturally arrive at different conclusions.

Finally, content analysis is frequently time-consuming and expensive. The actual task of examining and categorizing large volumes of content is often laborious and tedious. Plowing through 100 copies of the *New York Times* or 50 issues of *Newsweek* magazine involves large chunks of time and a corresponding degree of patience. In addition, if television content is selected for analysis, some means of preserving the programs for detailed examination is necessary. Typically, researchers videotape programs for analysis, but this requires access to a recorder and large supplies of videotape, materials not all researchers can afford.

Steps in Content Analysis

In general, a content analysis is conducted in several discrete stages. While they are listed here in sequence, the steps need not be followed in this order. In fact, the initial stages of analysis can easily be combined. Nonetheless, the following steps may be used as a rough outline:

1. Formulate the research question or hypothesis.
2. Define the population in question.
3. Select an appropriate sample from the population.
4. Select and define a unit of analysis.
5. Construct the categories of content to be analyzed.
6. Establish a quantification system.
7. Conduct a pilot study to establish reliability.
8. Code the content according to established definitions.
9. Analyze the collected data.
10. Draw conclusions and search for indications.

Formulating a Research Question

One problem to avoid in content analysis is the "counting-for-the-sake-of-counting" syndrome. The ultimate goal of the analysis must be clearly articulated in order to avoid aimless exercises in data collection that have little utility for mass media research. For example, it is possible to count the punctuation marks that are used in the *New York Times* and *Esquire* and generate statements such as, "*Esquire* used 45% more commas, but 23% fewer semicolons than the *New York Times.*" The value of such information for mass media theory or policy making, however, is dubious. Content analysis should not be conducted simply because the material exists and can be tabulated.

As with other methods of mass media research, content analyses should be guided by well-formulated research questions or hypotheses. A basic review of the literature is a required step. The sources for hypotheses are the same as for other areas of media research. It is possible to generate a research question based on existing theory, prior research, or practical problems, or as a response to changing social conditions. For example, a research question might ask if the growing visibility of the women's movement has produced a change in the way women are depicted in advertisements. Or, a content analysis might be conducted to determine if the public affairs programming of group-owned television stations differs from that of other stations. Well-defined research questions or hypotheses enable the development of accurate and sensitive content categories, which in turn helps to produce more valuable data.

Defining the Universe

This stage is not as grandiose as it sounds. To "define the universe" is to specify the boundaries of the body of content to be considered, which requires an appropriate operational definition of the relevant population. If researchers are interested in analyzing the content of popular songs, they must define what is meant by a "popular" song: All songs listed in *Billboard*'s Hot 100 chart? The top 50 songs? The top 10? What time period will be considered? The last six months? This month? If a researcher intends to study the image of minority groups on television, then he or she must define what "television" consists of. Is it evening programming, or does it also include daytime shows? Will the study examine news content or confine itself to dramatic offerings? By now, the requirements should be clear: what is needed is a concise statement that spells out the appropriate parameters of the investigation; for example:

> This study considers TV commercials broadcast in prime time in the New York City area from September 1, 1982, to October 1, 1982.

Or:

> This study considers the news content on the front pages of the *Washington Post* and the *New York Times*, excluding Sundays, from January 1 to December 31 of the past year.

Selecting a Sample

Once the population is defined, the next step is to select a sample. While many of the guidelines and procedures discussed in Chapter 4 are applicable here, the sampling of content involves some special considerations. On one hand, some analyses are concerned with a relatively finite amount of data, and it may be possible to conduct a census of the content. Thus, Wurtzel (1975) was able to perform a census of two years of public access television programming in New York, and Wimmer and Haynes (1978) conducted a census of seven years' worth of articles published in the *Journal of Broadcasting.* On the other hand, the more typical situation is one in which the researcher has such a vast amount of content available that a census is not practical. In this situation, a sample must be drawn.

Most content analysis in mass media involves multistage sampling. This process typically involves two stages (although it may entail three). The first stage is usually to take a sampling of content sources. For example, a researcher interested in the treatment of Watergate by American newspapers would first need to sample from among the 1,700 or so newspapers published each day. The researcher may decide that he is interested primarily in the way big city dailies covered the story and opt to analyze only the leading circulation newspapers in the ten largest American cities. To take another example, a researcher interested in the changing portrayal of the elderly in magazine advertisements would first need to sample from among the thousands of magazines available. In this instance, the researcher might select only the top 10, 15, or 25 mass circulation magazines. Of course, it is also possible to sample randomly if the task of analyzing all the titles is too overwhelming. A further possibility is to use the technique of stratified sampling discussed in Chapter 4. For example, the Watergate researcher might wish to stratify the sample by circulation size and sample from within strata composed of big city papers, medium city papers, and small city papers. The magazine researcher might stratify by type of magazine: news, women's interests, men's interests, and so on. A researcher interested in television content might stratify by network or by program type.

Once the sources are identified, the next sampling stage is the selection of dates. In many studies, the question of the time period from which the issues are to be selected will be determined by the goal of the project. If the goal is to assess the nature of Watergate news, then the sampling period is fairly well defined by the actual duration of the story. If the research question is directed toward changes in the image of women after the introduction of the equal rights amendment, then content should be sampled before, during, and after the amendment. But within this period, what editions of the newspaper, magazine, or television program will be selected for analysis? It would be a tremendous amount of work to analyze each copy of *Time, Newsweek,* and *U.S. News and World Report* over a five-year period. It is possible to sample from within that time period and obtain a representative group of issues. A simple random sample of the calendar dates involved is one possibility: after a random start, every *n*th issue of a publication is selected for the sample. This method cannot be used without planning, however, For instance, if the goal is 50 edition dates, and an interval of 7 is used, the sample may include 50 Saturday editions (periodicity). Since news content is not randomly distributed over the days of the week, the sample will not be representative.

Another technique for sampling edition dates is to stratify by week of the month and by day of the week. A sampling rule that no more than two days from one week can be chosen is one way to ensure a balanced distribution across the month. Another procedure is to construct a composite week for each month in the sample. For example, a study might use a sample of one Monday, drawn at random from the four or five possible Mondays in the month; one Tuesday drawn from the available Tuesdays; and so on, until all weekdays are included. How many edition dates should be selected? Obviously, this depends on the topic under study. If an investigator is trying to describe the portrayal of Mexican-Americans on prime time television, a large number of dates would have to be sampled in order to conduct a representative analysis. If there is an interest in analyzing the geographic sources of news stories, a smaller number of dates would be needed, since almost every story would be relevant to the analysis. The number of dates should be a function of the incidence of the phenomenon in question: the lower the incidence, the more dates will have to be sampled.

There are some rough guidelines for sampling in the media. Stempel (1952) drew separate samples of 6, 12, 18, 24, and 48 issues of a newspaper and compared the average content of each of the sample sizes in a single subject category against the total for the entire year. He found that each of the 5 sample sizes was adequate and that increasing the sample beyond 12 issues did not significantly improve upon accuracy. In television, Gerbner and his associates (1977) demonstrated that, at least for the purpose of measuring violent behavior, a sample of one week of fall programming will produce results comparable to various sample dates drawn throughout the year. As a general rule, however, the larger the sample the better, within reason, of course. If too few dates are selected for analysis, the possibility of an unrepresentative sample is increased. Larger samples, if chosen randomly, usually run less risk of being atypical.

Another problem to examine during the sampling phase relates to systematic bias in the content itself. For example, a study of the amount of sports news in a daily paper might yield inflated results if the sampling were done only in the month of April, when three or more professional sports are simultaneously in season. A study of marriage announcements in the Sunday *New York Times* for the month of June from 1932 to 1942 revealed no announcement of a marriage in a synagogue (Hatch and Hatch 1947). It was later pointed out that the month of June usually falls within a period during which traditional Jewish marriages are prohibited. Researchers familiar with their topics can generally detect and guard against this type of distortion.

Once the sources and the dates have been determined, there may be one further stage of sampling. A researcher might wish to

Figure 8.1 Hypothetical example of multistage sampling in content analysis

Research Question: Have there been changes in the type of products advertised in men's magazines from 1971 to 1980?

Sampling Stage One: Selection of Titles
Men's magazines are defined as those magazines whose circulation figures show that 80% or more of their readers are men. These magazines will be divided into two groups: large and medium circulation.

Large circulation: Reaches more than 1,000,000 men.
Medium circulation: Reaches between 500,000 and 999,999 men.

From all the magazines that fall into these two groups, three will be selected at random from each division, for a total of six titles.

Sampling Stage Two: Selection of Dates
Three issues from each year will be chosen at random from clusters of four months. One magazine will be selected from the January, February, March, April issues, etc. This procedure will be followed for each magazine, yielding a final sample of 30 issues per magazine, or 180 total issues.

Sampling Stage Three: Selection of Content
Every other display ad will be tabulated, regardless of size.

confine his or her attention to a selection of content within an edition. For example, an analysis of the front page of a newspaper is valid for a study of general reporting trends but is probably inadequate for a study of social news coverage.

Figure 8.1 provides an example of multistage sampling in content analysis.

Selecting the Unit of Analysis

The **unit of analysis** is the thing that is actually counted. It is the smallest element of a content analysis, but it is also one of the most important. In written content, the unit of analysis might be a single word or symbol, a theme (a single assertion about one subject), or an entire article or story. In television and film analyses, units of analysis can be characters, acts, or entire programs. Specific rules and definitions are required for determining these units to allow for greater agreement between coders and fewer "judgment calls."

Certain units of analysis are simpler to count than others. It is easier to determine the number of stories on the "CBS Evening News" that deal with international news than the number of acts of violence in a week of network television because a "story" is a more readily distinguishable unit of analysis than an "act." The beginning and ending of a news story are fairly easy to see, but what would occur if the researcher trying to catalog violent content was faced with a long fistfight between three characters? Is that one act of violence? Is every blow considered an act? What if another character joins in? Does it then become a different act?

Figure 8.2 Examples of units of analysis

Study	*Title*	*Unit*
Dominick and Rauch (1972)	"The Image of Women in Network TV Commercials"	Female characters who fulfilled either of the following criteria: (1) appeared on screen for at least three seconds, or (2) had at least one line of dialogue
Wurtzel (1975)	"Public Access Cable Television: Content"	All programs cablecast on the New York public access channels between July 1971 and June 1973
Greenberg *et al.* (1980)	"Three Seasons of Television Characters"	Each speaking character
Miller (1975)	"The Content of News Photos: Women's and Men's Roles"	Every man and woman appearing in a news photo except those less than 13 years old, those in standing shots (columnists), those in crowd shots, and those appearing in photos in special sections
Baker and Walter (1975)	"The Press as a Source of Information About Activities of a State Legislature"	All reports on the legislature longer than one sentence

Operational definitions of the unit of analysis should be clear-cut and thorough; the criteria for inclusion should be apparent and easily observed. These goals cannot be accomplished without effort and some trial and error. As a preliminary step, researchers must form a rough draft of a definition and then sample representative content to see if problems exist. This procedure usually results in further refinement and modification of the operational definition. For illustration, Figure 8.2 contains typical operational definitions of units of analysis taken from mass media research.

Constructing Categories for Analysis

At the heart of any content analysis is the category system used to classify media content. The precise makeup of this system, of course, varies with the topic under study. As Berelson (1952) pointed out, "Particular studies have been productive to the extent that the categories were clearly formulated and well-adapted to the problem and the content."

To be serviceable, all category systems should be mutually exclusive, exhaustive, and reliable. A category system is mutually exclusive if a unit of analysis can be placed in one and only one category. If the researcher discovers that certain units fall simultaneously into two different categories, he or she must revise the definitions of those categories. For example, suppose researchers attempted to describe the ethnic makeup of prime time television characters using the following category system: (1) Black; (2) Jewish; (3) White; (4) American Indian; (5) other. As is obvious, a Jewish person would fall into two categories at once, thus violating the exclusivity rule. Or, to take another example, a researcher might start with the following categories in an attempt to describe the types of programs on network television: (1) situation comedies; (2) children's shows; (3) movies; (4) documentaries; (5) action/adventure programs; (6) quiz and talk shows; (7) general drama. This list might look acceptable at first glance, but a program such as "Little House on the Prairie" raises questions. Is it to be placed in the "children's shows" category or in "general drama"? Definitions must have a high degree of specificity to handle problems such as this.

In addition to exclusivity, content analysis categories must have the property of **exhaustivity:** every unit of analysis should be able to be placed into an existing slot. If investigators suddenly find a unit of analysis that does not logically fit into a predefined category, they have a problem with their category system. Taken as a whole, the category system should account for every unit of analysis. Achieving exhaustivity is usually not a great problem in mass media content analysis. If one or two unusual instances are detected, a category labeled "other" or "miscellaneous" usually solves the problem. (If too many items fall into this category, however, a reexmination of the original category definitions is called for; a study with 10% or more of its content in the "other" category is probably overlooking some relevant content characteristic). An additional way to assure exhaustivity is to dichotomize or trichotomize the content: attempts at problem-solving might be defined as aggressive and non-aggressive; statements might be placed in positive-neutral-negative categories. The most practical way to determine if a proposed categorization system is exhaustive is to pretest it on a sample of content. If unanticipated items appear, the original scheme requires changes before the primary analysis can begin.

The categorization system should also be reliable; that is, different coders should agree in the great majority of instances about the proper category for each unit of analysis. This agreement is usually quantified in content analysis and is called **intercoder reliability.** Precise category definitions generally increase reliability, while sloppily defined categories tend to lower it. Pretesting the category system for reliability is highy recommended before actually beginning to process the main body of content. Reliability is crucial in content analysis and is discussed in more detail later in this chapter.

A question that may arise in constructing category systems is how many categories to include. Common sense, pretesting, and practice with the coding system are valuable guides to aid the researcher in steering between the two extremes: developing a system with too few categories, so that essential differences are obscured; and defining too many categories, so that only a small percentage falls into each, thus limiting generalizations. As an illustration of too few categories, consider Wurtzel's (1975) study of programming on public access television, where one of the preliminary categories was labelled "informational." The data indicated that more than 70% of the content fell into this classification. As a result, Wurtzel subdivided the category into seven different informational headings: ethnic, community, health, consumer, etc. An example of the opposite extreme is the attempt made by Dominick, Richman, and Wurtzel (1979) to describe the types of problems encountered by characters in prime time television shows popular with children. They originally developed seven different categories, among which were one for problems that dealt with romance, one for problems between friends, and one for other emotional problems arising out of relationships (such as siblings or co-workers). Preliminary analysis, however, indicated that only a small fraction of problems fell into the "friendship" and "other emotional" slots. Consequently, these three categories were combined into a single classification labeled "problems dealing with romance, sentiment, and other emotions." As a general rule, many researchers suggest that too many initial categories are preferable to too few, since it is usually easier to combine several categories than it is to subdivide one large one after the units have been codified.

Establishing a Quantification System

Quantification in content analysis can involve all four of the levels of data measurement discussed in Chapter 2, although usually only nominal, interval, and ratio data are used. At the nominal level, researchers simply count the frequency of occurrence of the units in each category. Gantz, Gartenberg, and Rainbow (1980) employed nominal measurement to determine the percentage of elderly people appearing in different categories of magazine advertisements: 14% in "corporate image" advertisements; 12% liquor; 11% travel; and so on. The topics of conversation on daytime television, the themes of newspaper editorials, and the occupations of prime time television characters can all be quantified by means of nominal measurement.

At the interval level, it is possible to develop scales for coders to use to rate certain attributes of characters or situations. For example, in a study dealing with the images of women in commercials, each character might be rated by coders on several scales such as:

independent ____:____:____:____:____ dependent

dominant ____:____:____:____:____ submissive

Scales such as these add depth and texture to a content analysis and are perhaps more interesting than the surface data obtained through nominal measurement. It should be noted, however, that rating scales inject subjectivity into the analysis and may thus lower intercoder reliability unless careful training is undertaken.

Ratio level measurements in mass media research generally relate to measurements of space and time. In the print media, column/inch measurements are used to analyze editorials, advertisements, and stories about particular events or phenomena. In television and radio, ratio level measurements are made concerning time: the number of commercial minutes, the types of programs on the air and the amount of the program day devoted to various types of programs, and so on.

Coding the Content

Placing a unit of analysis into a content category is called **coding.** It is the most time-consuming and least glamorous part of a content analysis. Individuals who do the coding are called *coders*. The number of coders involved in a content analysis is typically small; a brief examination of a sampling of recent content analyses indicates that from three to six coders are typically used. All coders should undergo one or more training sessions during which they are familiarized with the definitions and units of analysis and are allowed to practice coding actual content. Careful training of coders usually results in a more reliable analysis.

In order to facilitate coding, standardized sheets are usually used (an example is shown below). These sheets allow coders to classify the data by simply placing check marks or slashes in predetermined spaces. If data are to be tabulated by hand, the coding sheets should be constructed to allow for rapid tabulation. Some studies code data on 4″ × 6″ index cards, with information recorded across the top of the card. This enables researchers to sort the information quickly into appropriate categories. Templates are available to speed the measurement of newspaper space. Researchers who work with television generally videotape the programs and allow coders to stop and start the tape at their own pace while coding data. If a tape machine is not available, the coding procedure has to be simplified, since the coders will not have the opportunity for a second look.

When a computer is used in tabulating data, the data are usually transferred directly to computer coding sheets or perhaps mark-sense forms or optical scan sheets (answer sheets scored by computer). These forms save time and reduce data errors.

Computers are useful not only in the data tabulation phase of a content analysis, but also in the actual coding process. Since computers can do simple tasks with extreme speed, it is possible for them to recognize words or even syllables as they occur in a sample of text. Wilhoit and Sherrill (1968), for example, instructed a computer to recognize the names of U.S. Senators as they appeared in wire service copy. Of course, the drawback to this approach is that the original

Example of Standardized Coding Sheets

Coder Initials ______ Program Code ______

PROBLEM SOLVING ATTEMPT DESCRIPTION SHEET

1. Program code number 01 02 03 04 06 07
2. Name of character attempting solution ______
(NOTE: ANSWER QUESTIONS 3-8 ONLY IF INFORMATION HAS NOT BEEN PREVIOUSLY RECORDED ON CHARACTER DESCRIPTION SHEET)

3. Is character
   ______ hero
   ______ villain
4. Is character
   ______ human
   ______ animal
   ______ other (specify)
5. Is character
   ______ male
   ______ female
6. Is character
   ______ kid (2-11)
   ______ teen (12-19)
   ______ young adult (20-49)
   ______ mature adult (50+)
7. Does character have magical or unrealistic powers?
   ______ yes ______ no

body of copy to be scanned must be in computer-readable form. In many instances, transcribing the copy into the necessary format would require far more effort than counting the data manually. This situation may change, however, with the advent of teletext and view-data systems where the copy is already machine-readable. In any case, the use of computers in content analysis is now widely accepted. For more information, the reader should consult Gerbner et al. (1969).

Analyzing the Data

The descriptive statistics discussed in Chapters 9–12, such as percentages, means, modes, and medians, are appropriate for content analysis. If hypothesis tests are planned, common inferential statistics (results are generalized to the population) are acceptable. The chi-square test is the most commonly used, since content analysis data tend to be nominal in form; however, if the data meet the requirements of interval or ratio levels, a t-test, ANOVA, or Pearson's *r* may be appropriate. Other statistical analyses are discussed by Krippendorf (1980), such as discriminant analysis, cluster analysis, and contextual analysis.

Interpreting the Results

If an investigator is testing specific hypotheses concerning the relationships between variables, the interpretation will be fairly evident. However, if the study is descriptive in nature, questions may arise as to the meaning or importance of the results. Researchers are often faced with a "fully-only" dilemma. Suppose, for example, that a content analysis of children's television programs reveals that 30% of the commercials in these programs were for snacks and candy. What is the researcher to conclude? Is this a high or a low amount? Should the researcher report, "*Fully* 30% of the commercials fell into this category," or should he or she present this same percentage in a different light: "*Only* 30% of the commercials fell into this category"? The point is that the investigator needs some benchmark for comparison. Thirty percent may indeed be a high figure when compared to commercials for other products or those shown during adult programs.

In a study done by one of the authors, the amount of network news time devoted to the various states was tabulated. When the counting was done, it was determined that California and New York receive 19% and 18%, respectively, of non-Washington, D.C. national news coverage. By themselves, these numbers are interesting, but their significance is somewhat unclear. In an attempt to aid interpretation, each state's relative news time was compared to its population, and an "attention index" was created by subtracting the ratio of each state's population to the national population from its percentage of news coverage. This provided a listing of states that were either "over-covered" or "under-covered" (Dominick 1977).

Reliability

The concept of reliability is crucial to content analysis. If a content analysis is to be objective, then its measures and procedures must be reliable. Reliability means that repeated measurement of the same material will result in similar decisions or conclusions. If the results fail to achieve reliability, then something is amiss with the coders, the coding instructions, the category definitions, the unit of analysis, or some combination of these. In order to achieve acceptable levels of reliability, the following steps are recommended:

1. Define category boundaries with maximum detail. A group of vague or ambiguously defined categories makes reliability extremely difficult to achieve. Examples of units of analysis and a brief explanation for each are necessary for coders to fully understand the procedure.

2. Train coders in using the coding instrument and the category system. Training sessions conducted before the data are collected can help to eliminate methodological problems. During these sessions, the group as a whole should code sample material; afterwards, they should discuss the results as well as the purpose of the study. Disagreements should be analyzed as they occur. The end result of the training sessions is a "bible" of detailed instructions and coding examples, of which each coder should receive a copy.

Figure 8.3 Example of poorly defined categories

Coders	Categories			
	I	II	III	IV
A	1,3,7,11	2,5,6,8,12,13	10	4,9,14
B	1,3,7,11	5,8,10,12	2,6,13	4,9,14
C	1,3,7,11	2,8,12,13	5,6,10	4,9,14
D	1,3,7,11	5,6	2,8,10,12,13,14	4,9

(Arabic numerals refer to items.)

3. Conduct a pilot study. Select a subsample of the content universe under consideration and let independent coders categorize it. These data are useful for two reasons: (1) poorly defined categories can be detected; and (2) chronically dissenting coders can be identified. To illustrate these problems, consider Figures 8.3 and 8.4.

In Figure 8.3, the definitions for Categories I and IV appear satisfactory. All four coders placed Units 1, 3, 7, and 11 in the first category; in Category IV, Item 14 is classified consistently by three of the four coders and Items 4 and 9 by all four coders. The confusion apparently lies in the boundaries between Categories II and III. Two coders put Items 2, 6, and 13 in Category II and two placed them in Category III. The definitions of these two categories require reexamination and perhaps revision in order to clarify this ambiguity.

Figure 8.4 illustrates the problem of the chronic disagreer.

Coders A and B agree with one another seven times out of eight. Coders B and C, however, agree only two times out of eight, and coders A and C agree only once. Obviously, Coder C is going to be a

Figure 8.4 Example of the chronic dissenter

Items	Coders		
	A	B	C
1	I	I	II
2	III	III	I
3	II	II	II
4	IV	IV	III
5	I	II	II
6	IV	IV	I
7	I	I	III
8	II	II	I

(Roman numerals refer to categories.)

problem. In such an instance, the investigator can carefully reexplain to this coder the rules used in categorization and examine the reasons for his or her consistent deviation. If the problem persists, it may be necessary to dismiss the coder from the analysis.

Assuming that the initial test of reliability yields satisfactory results, the main body of data is then coded. When the coding is complete, it is recommended that a subsample of the data, probably between 10% and 25%, be reanalyzed by independent coders in order to calculate an overall intercoder reliability coefficient.

Intercoder reliability can be calculated by several methods. Holsti (1969) reported a formula for determining the reliability of nominal data in terms of percentage of agreement:

$$\text{Reliability} = \frac{2M}{N_1 + N_2}$$

Where M is the number of coding decisions on which two coders agree, and N_1 and N_2 refer to the total number of coding decisions by the first and second coder, respectively

Thus, if two coders judge a subsample of 50 units and agree on 35 of them, the calculation is:

$$\text{Reliability} = \frac{2(35)}{50 + 50} = .70$$

While this method is straightforward and easy to calculate, it is criticized because it does not take into account the fact that some coder agreement occurs strictly by chance, and that this amount is a function of the number of categories in the analysis. For example, a two-category system should obtain 50% reliability simply by chance; a five-category system would generate a 20% agreement by chance; and so on. To take this into account, Scott (1955) developed the *pi* index, which corrects for the number of categories used and also for the probable frequency of use.

$$\text{pi} = \frac{\%\text{ observed agreement} - \%\text{ expected agreement}}{1 - \%\text{ expected agreement}}$$

A hypothetical example will demonstrate this index. Suppose that two coders are assigning magazine advertisements to the following six categories and obtain the following distribution:

Category	*Percent of All Ads*
1 Prestige appeal	30%
2 Economic appeal	20%
3 Affiliation appeal	20%
4 Fear appeal	15%
5 Utilitarian appeal	10%
6 Other appeal	5%

First, the percentage of expected agreement must be calculated. This is the sum of the squared percentages of all categories. Thus, % expected agreement = $(.3)^2 + (.2)^2 + (.2)^2 + (.15)^2 + (.1)^2 + (.05)^2 =$ .20. If the coders agree on 90% of their classifications (% observed agreement), then *pi* can be calculated as follows:

$$\text{pi} = \frac{.90 - .20}{1 - .20} = .875$$

Fless (1971) extended this notion to situations where the same content is seen simultaneously by a number of coders.

Estimating reliability with interval data requires certain care. Several studies have used the correlation method called the Pearson *r*, a method that investigates the relationship between two items. The Pearson *r* can range from −1.00 and +1.00. In estimating reliability in content analysis, however, if this measure has a high value, it may indicate one of two different things: that the coders were in agreement, or that their ratings were associated in some systematic manner.

For example, suppose an interval scale ranging from 1 to 10 is used to score the degree of favorability of a news item toward some person or topic: a score of "1" represents very positive, while "10" represents very negative. Assume that two coders are independently scoring the same ten items. Figure 8.5 illustrates two possible outcomes. In Situation I, the coders agree on every item, and *r* equals 1.00. In Situation II, the coders disagree on every item by three scale positions, yet *r* still equals 1.00. Clearly, the two situations are not equally reliable.

Krippendorf (1980) circumvents this dilemma by presenting what might be termed an "all-purpose reliability measure," *alpha*, which can be used for nominal, ordinal, interval, and ratio scales and

Figure 8.5 Problems with using *r* as a reliability measure

Situation I			Situation II		
Items	Coder 1	Coder 2	Items	Coder 1	Coder 2
1	1	1	1	1	4
2	2	2	2	2	5
3	3	3	3	3	6
4	3	3	4	3	6
5	4	4	5	4	7
6	5	5	6	5	8
7	6	6	7	6	9
8	6	6	8	6	9
9	7	7	9	7	10
10	7	7	10	7	10
	$r = 1.00$			$r = 1.00$	

for more than one coder. Although somewhat difficult to calculate, *alpha* is the equivalent of Scott's *pi* at the nominal level with two coders and represents an improvement over *r* in the interval situation.

What is an acceptable level of intercoder reliability? This question depends on the research context and the type of information coded. There are instances where little coder judgment is needed to place units into categories and coding becomes a mechanical or clerical task, such as counting the number of words per sentence in a newspaper story or tabulating the number of times a network correspondent contributes a story to the evening news. In this context, one would expect a fairly high degree of reliability, perhaps approaching 100%, since coder disagreements would probably be the result only of carelessness and fatigue. If, however, a certain amount of interpretation is involved, reliability estimates are typically lower. In general, the greater the amount of judgmental leeway given to coders, the lower the reliability coefficients will be. As a rule of thumb, most published content analyses typically report a minimum reliability coefficient of about 90% or above when using Holsti's formula, and about .75 or above when using *pi* or Krippendorf's *alpha*.

Validity

In addition to being reliable, a content analysis must yield valid results. As indicated in Chapter 2, validity is usually defined as the degree to which an instrument actually measures what it sets out to measure. This raises special concerns in content analysis. In the first place, validity is intimately connected with the procedures used in the analysis. If the sampling design is faulty, if categories overlap, or if reliability is low, then the results of the study probably possess little validity.

Additionally, the adequacy of the definitions used in a content analysis bears directly on the question of validity. For example, a great deal of content analysis has focused on the depiction of televised violence; different investigators have offered different definitions of what constitutes a violent act. The question of validity emerges when one tries to decide whether each of the various definitions actually encompasses what one might logically refer to as violence. The continuing debate between Gerbner and the television networks vividly illustrates this problem. Gerbner *et al.*'s (1977) definition of violence includes accidents, acts of nature, or violence that might occur in a fantasy or a humorous setting. However, network analysts do not consider these to be acts of violence (Blank 1977). Both Gerbner and the networks offer arguments in support of their decisions. Which analysis is more valid? The answer depends in part on the plausibility of the rationale that underlies the definitions.

This discussion relates closely to a technique traditionally called *face validity*. This validation technique assumes that an instrument adequately measures what it purports to if the categories are rigidly and satisfactorily defined and if the procedures of the analysis have been adequately conducted. Most descriptive content analyses usually rely upon face validity, but there are other techniques available.

The use of *concurrent validity* in content analysis is exemplified in a study by Clarke and Blankenburg (1972). These investigators attempted a longitudinal study of violence in television shows dating back to 1952. Unfortunately, few copies of the early programs were available, and the authors were forced to rely on program summaries in *TV Guide*. In order to establish that such summaries would disclose the presence of violence, the authors compared the results of a subsample of programs coded from these synopses to the results obtained from a direct viewing of the same programs. These results were sufficiently related to convince the authors that their measurement technique was valid. However, this method of checking validity is only as good as the criterion measurement: if the direct viewing technique is itself invalid, there is little value in showing that synopsis coding is related to it.

Only a few studies have attempted to document *construct validity*. One instance involves the use of "sensationalism" in news stories. This construct has been measured by semantic differentials and factor analysis (see Chapter 12), in an attempt to isolate its underlying dimensions, and related to relevant message characteristics (Tannenbaum and Lynch 1960; Tannenbaum 1962). Another technique that investigators occasionally use is *predictive validity*. For example, certain content attributes from wire stories might allow a researcher to predict which items will be carried by a newspaper and which will not.

In summary, several different methods to assess validity are used in content analysis. The most common is face validity. While this may be appropriate for some studies, it is recommended that the content analyst also examine other methods to establish the validity of a given study.

Examples of Content Analysis

Figure 8.6 summarizes four recent content analyses. The purpose of the analysis, the sample, the unit of analysis, illustrative categories, and the type of statistic are listed for each study. The complete citation for each of the studies is contained at the end of the chapter.

Summary

Content analysis is a popular research technique in mass media. Many of the steps involved in laboratory and survey studies are also

Figure 8.6 Examples of content analysis

Study	*Purpose*	*Sample*	*Unit of Analysis*	*Representative Categories*	*Statistic*
Fine (1981)	To examine the relationships between soap opera conversation and real life conversation	20 randomly selected episodes of four soap operas	Each dyadic conversation	Topics of conversation: marriage, family, romance, health, friends, etc.	Crosstabs Chi-square
Breed and DeFoe (1981)	To examine the motives, contents, and outcomes of drinking behavior on television	225 episodes of 30 prime time situation comedies and dramas	Any character involved in a scene that portrayed alcohol consumption	Consequences of drinking: strained relationships, danger to self, harm to self, and so on	Crosstabs Chi-square
Singletary (1978)	To identify changes in front-page news photos from 1936 to 1976	Front-page photos appearing in one week of issues of six urban papers for each month of the sample years	All front-page photos except mug shots	Content of photos: crime, sex, science, sports, health, violence, etc.	Differences between proportions
Sparkes (1978)	To describe the flow of news between the U.S. and Canada	Two composite weeks of issues from 10 Canadian papers and 29 U.S. papers drawn from the first six months of 1975	All news items in each issue of each paper	Sources of news: staff, AP, UPI, CP, Reuters, etc.	Percentages *r* correlation

found in content analysis; in particular, sampling procedures need to be objective and detailed, and operational definitions are mandatory. Interpreting a content analysis, however, requires more caution: no claims about the impact of the content can be drawn from an analysis of the message without an accompanying study that examines the audience.

Questions and Problems for Further Investigation

1. Define a unit of analysis that could be used in a content analysis of:
   a. Problem-solving on television
   b. News emphasis in a daily and a weekly newspaper
   c. Changes in the values expressed by popular songs
   d. The role of women in editorial cartoons
2. Using the topics in Question 1, define a sample selection procedure that would be appropriate for each.
3. Generate two content analyses that could be used as preliminary tests for an audience study.
4. Conduct a brief content analysis of one of the topics listed below. (Train a second individual in the use of the category system that you develop and have this person independently code a subsample of the content.)
   a. Similarities and differences between local newscasts on two television stations
   b. Changes in the subject matter of movies from 1970 to 1980
   c. The treatment of the elderly on network television
5. Using the topic selected in Question 4, compute a reliability coefficient for the items that were scored by both coders.

References and Suggested Readings

Adams, W., and Ferber, P. 1977. "Television Interview Shows: The Politics of Visibility." *Journal of Broadcasting* 21:141–51.

Adams, W., and Schreibman, F. 1978. *Television Network News: Issues in Content Research.* Washington, D.C.: George Washington University.

Baker, K., and Walter, B. 1975. "The Press as a Source of Information about Activities of a State Legislature." *Journalism Quarterly* 52:735–40.

Blank, D. 1977. "The Gerbner Violence Profile." *Journal of Broadcasting* 21: 273–79.

Berelson, B. 1952. *Content Analysis in Communication Research.* New York: The Free Press.

Breed, W., and DeFoe, J. 1981. "The Portrayal of the Drinking Process on Prime Time Television." *Journal of Communication* 31:58–67.

Clarke, D., and Blankenburg, W. 1972. "Trends in Violent Content in Selected Mass Media." *Television and Social Behavior: Media Content and Control,* ed. by G. Comstock and E. Rubinstein. Washington, D.C.: U. S. Government Printing Office.

Comstock, G. 1975. *Television and Human Behavior: The Key Studies*. Santa Monica: Rand Corporation.

Davis, F. 1951. "Crime News in Colorado Newspapers." *American Journal of Sociology* 57:325–30.

DeFleur, M. 1964. "Occupational Roles as Portrayed on Television." *Public Opinion Quarterly* 28:57–74.

DeFleur, M., and DeFleur, L. 1967. "The Relative Contribution of Television as a Learning Source for Children's Occupational Knowledge." *American Sociological Review* 32:777–89.

Dominick, J., and Rauch, G. 1972. "The Image of Women in Network TV Commercials." *Journal of Broadcasting* 16:259–65.

Dominick, J. 1973. "Crime and Law Enforcement on Prime Time Television." *Public Opinion Quarterly* 37:241–50.

Dominick, J., Wurtzel, A., and Lometti, G. 1975. "Television Journalism vs. Show Business: A Content Analysis of Eyewitness News." *Journalism Quarterly* 52:213–18.

Dominick, J. 1977. "Geographic Bias in National TV News." *Journal of Communication* 27:94–99.

Dominick, J., Richman, S., and Wurtzel, A. 1979. "Problem-Solving in TV Shows Popular with Children: Assertion vs. Aggression." *Journalism Quarterly* 56:455–63.

Fine, M. 1981. "Soap Opera Conversations." *Journal of Communication* 31: 97–107.

Fless, J. L. 1971. "Measuring Nominal Scale Agreement Among Many Raters," *Psychological Bulletin* 76:378–82.

Gantz, W., Gartenberg, H., and Rainbow, C. 1980. "Approaching Invisibility: The Portrayal of the Elderly in Magazine Advertisements." *Journal of Communication* 30:56–60.

Gerbner, G., Holsti, O., Krippendorf, K., Paisley, W., and Stone, P. 1969. *The Analysis of Communication Content*. New York: Wiley.

Gerbner, G. 1969. "The Television World of Violence." *Mass Media and Violence*, ed. by D. Lange, R. Baker, and S. Ball. Washington, D.C.: U. S. Government Printing Office.

Gerbner, G., Gross, L., Jackson-Beeck, M., Jeffries-Fox, S., and Signorielli, N. 1977. "One More Time: An Analysis of the CBS 'Final Comments on the Violence Profile.' " *Journal of Broadcasting* 21:297–304.

Gerbner, G., Gross, L., Signorielli, N., Morgan, M., and Jackson-Beeck, M. 1979. "The Demonstration of Power: Violence Profile No. 10." *Journal of Communication* 29:177–98.

Greenberg, B., and Kahn, S. 1970. "Blacks in *Playboy* Cartoons." *Journalism Quarterly* 47:557–60.

Greenberg, B., Simmons, K., Hogan, L., and Atkin, C. 1980. "Three Seasons of Television Characters: A Demographic Analysis." *Journal of Broadcasting* 24:49–60.

Hatch, D., and Hatch, M. 1947. "Criteria of Social Status as Derived from Marriage Announcements in the *New York Times.*" *American Sociological Review* 12:396–403.

Hennessee, J., and Nicholson, J. 1972. "NOW Says: TV Commercials Insult Women." *New York Times Magazine* May 28, 1972, 12–14.

Holsti, O. 1969. *Content Analysis for the Social Sciences and Humanities.* Reading, Mass.: Addison-Wesley.

Katzman, N. 1972. "Television Soap Operas." *Public Opinion Quarterly* 36: 200–12.

Kerlinger, F. 1973. *Foundations of Behavioral Research.* 2nd ed. New York: Holt, Rinehart and Winston.

Krippendorf, K. 1980. *Content Analysis: An Introduction to Its Methodology.* Beverly Hills: Sage.

Merritt, S., and Gross, H. 1978. "Women's Page/Lifestyle Editors: Does Sex Make a Difference?" *Journalism Quarterly* 55:508–14.

Mertz, R. 1970. "Analysis of the Portrayal of Older Americans in Commercial Television Programming." Paper presented to the International Communication Association, Minneapolis.

Miller, S. 1975. "The Content of News Photos: Women's and Men's Roles." *Journalism Quarterly* 52:70–75.

Scott, W. 1955. "Reliability of Content Analysis: The Case of Nominal Scale Coding." *Public Opinion Quarterly* 17:321–25.

Singletary, M. 1978. "Newspaper Photographs A Content Analysis, 1936–76." *Journalism Quarterly* 55:585–89.

Sparkes, V. 1978. "The Flow of News Between Canada and the United States." *Journalism Quarterly* 55:260–68.

Stempel, G. 1952. "Sample Size for Classifying Subject Matter in Dailies." *Journalism Quarterly* 29:333–34.

Tannenbaum, P., and Lynch, M. 1960. "Sensationalism: The Concept and Its Measurement." *Journalism Quarterly* 37:381–92.

Tannenbaum, P. 1962. "Sensationalism: Some Objective Message Correlates." *Journalism Quarterly* 39:317–23.

Tannenbaum, P., and Greenberg, B. 1968. "Mass Communication." *Annual Review of Psychology* Vol. 19, 351–86.

Walizer, M. H., and Wienir, P. L. 1978. *Research Methods and Analysis.* New York: Harper & Row.

Wilhoit, G., and Sherill, K. 1968. "Wire Service Visibility of U.S. Senators." *Journalism Quarterly* 45:42–48.

Wimmer, R., and Haynes, R. 1978. "Statistical Analyses in the *Journal of Broadcasting.*" *Journal of Broadcasting* 22:241–48.

Wurtzel, A. 1975. "Public Access Cable Television: Programming." *Journal of Communication* 25:15–21.

P A R T 3

Basic Statistics

CHAPTER 9

Introduction to Statistics

Statistics provide mathematical methods for organizing, summarizing, and analyzing data that have been collected and measured. Statistics cannot perform miracles, however. If a research question is misdirected, poorly phrased, or ambiguous, or if a study uses sloppy measurement and design, a statistical technique cannot magically salvage the project. Statistics can produce valid results only if the data variables they comprise have been established scientifically.

Part Three focuses on the statistical procedures used by mass media researchers. This chapter provides an introduction to the two types of research statistics: descriptive and inferential.

Descriptive Statistics

Descriptive statistics are intended to reduce data sets to allow for easier interpretation. Suppose an investigator asks 100 people how long they listened to radio yesterday and records all 100 answers at random on a sheet of paper. He or she would be hard pressed to draw conclusions from a simple examination of that paper. Analysis of the data would be facilitated if they were organized in some fashion. In this regard, descriptive statistics can be useful.

During the course of a research study, investigators typically collect data that are the results of measurements or observations of the people or items in the sample. These data usually have little meaning or usefulness until they are displayed or summarized using one of the techniques of descriptive statistics. There are two primary methods by which mass media researchers can make their data more manageable: data distribution and summary statistics.

Data Distribution

One way researchers can display their data is by distributing it in tables or graphs. A **distribution** is simply a collection of numbers. Table 9.1 is a hypothetical distribution of the responses of 20 people to the question, "How many hours did you listen to the radio yesterday?"

It is difficult for the investigator to draw any conclusions or make any generalizations from this collection of unordered scores.

As a preliminary step toward making these numbers more manageable, the data may be arranged in a **frequency distribution**; that is, a table of each possible score, ordered according to magnitude, and its actual frequency of occurrence. Table 9.2 presents the data from the hypothetical survey in a frequency distribution.

Now the data begin to show more of a pattern. Note that the typical frequency distribution table consists of two columns. The first, on the left, contains all the possible values of the variable under study, and the second, on the right, shows the number of occurrences of each value. The sum of the frequency column is the number (N) of persons or items that make up the distribution.

Table 9.1 Distribution of Responses to "How Many Hours Did You Listen to the Radio Yesterday?"

Respondent		Respondent	
A	0.0	K	0.0
B	2.5	L	2.0
C	1.0	M	1.0
D	2.0	N	3.0
E	0.0	O	0.5
F	0.5	P	0.5
G	1.0	Q	1.0
H	1.0	R	1.0
I	1.5	S	1.0
J	0.5	T	0.0

Table 9.2 Frequency Distribution of Responses to "How Many Hours Did You Listen to the Radio Yesterday?"

Hours	Frequency
0.0	4
0.5	4
1.0	7
1.5	1
2.0	2
2.5	1
3.0	1

A frequency distribution can also be constructed using grouped intervals, each of which contains several score levels. Table 9.3 shows the data from the "radio listening" survey with the scores grouped together in intervals. Note that while Table 9.3 is a more compact frequency distribution than Table 9.2, the scores have lost their individual identity.

Other columns can be included in frequency distribution tables. For example, the data can be transformed into proportions or percentages. To obtain the percentage of a response, simply divide the frequency of the response by N, or the total number of responses in the distribution. Percentages allow comparisons to be made between frequency distributions that are based on different Ns.

Another column that appears in some frequency distributions gives the cumulative frequency (cf); it is constructed by adding the number of scores in one interval to the number of scores in the intervals above it.

Table 9.4 displays the frequency distribution from Table 9.2 with the addition of a percentage column, a cumulative frequency column, and a column showing cumulative frequency as a percentage of N.

Sometimes it is desirable to present data in graph form. The graphs described below contain the same amount of information as frequency distributions. Graphs usually consist of two perpendicular lines, the *x-axis* (horizontal) and the *y-axis* (vertical). Over the years, statisticians have developed certain conventions regarding graphic format. One common convention is to list the scores along the x-axis and the frequency or relative frequency along the y-axis. Thus, the height of a line or bar indicates the frequency of a score.

Table 9.3 Frequency Distribution of Radio Listening Scores Grouped in Intervals

Hours	Frequency
0.00 – 0.50	8
0.51 – 1.50	8
1.51 – 3.00	4

Table 9.4 Frequency Distribution of Radio Listening Scores

Hours	Frequency	Percentage	cf	cf Percentage
0.0	4	20%	4	20%
0.5	4	20%	8	40%
1.0	7	35%	15	75%
1.5	1	5%	16	80%
2.0	2	10%	18	90%
2.5	1	5%	19	95%
3.0	1	5%	20	100%
	N = 20	100%		

One common form of graph is the **histogram,** or bar chart, in which frequencies are represented by vertical bars. Figure 9.1 is a histogram constructed from the data in Table 9.1. Note that the scores across the x-axis are actually the midpoints of half-hour intervals: the first category extends from 0 hours to .2499 hours, the second from .25 hours to .7499 hours, and so on.

If a line is drawn from the midpoint of each interval at its peak along the y-axis to each adjacent midpoint/peak, the resulting graph is called a **frequency polygon.** Figure 9.2 shows a frequency polygon superimposed onto the histogram from Figure 9.1. As can be seen, the two figures display the same information.

A **frequency curve** is similar to a frequency polygon except that points are connected by a continuous, unbroken curve instead of lines. Such a curve assumes that any irregularities shown in a frequency polygon are simply due to chance and that the variable being

Figure 9.1 Histogram of radio listening scores

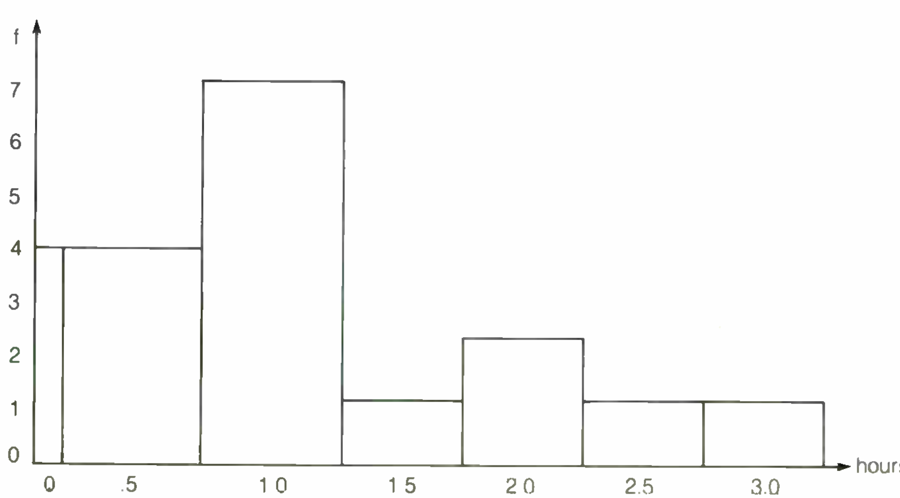
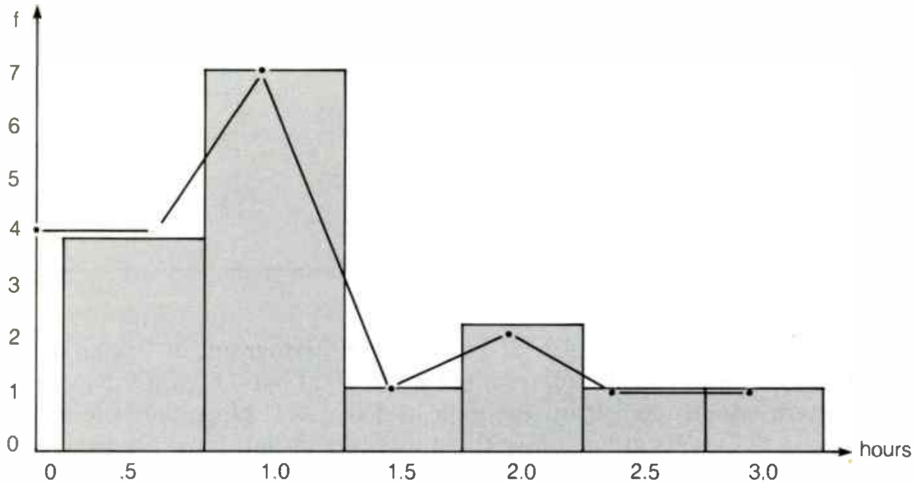


Figure 9.2 Frequency polygon of radio listening scores



studied is distributed continuously over the population. Figure 9.3 superimposes a frequency curve onto the frequency polygon shown in Figure 9.2.

Frequency curves are described in relation to the **normal curve,** a symmetrical bell curve whose properties are discussed more fully later in this chapter. Figure 9.4 illustrates the normal curve and shows the ways in which a frequency curve can deviate from it. These patterns of deviation are referred to as skewness and kurtosis.

Figure 9.3 Frequency curve of radio listening scores

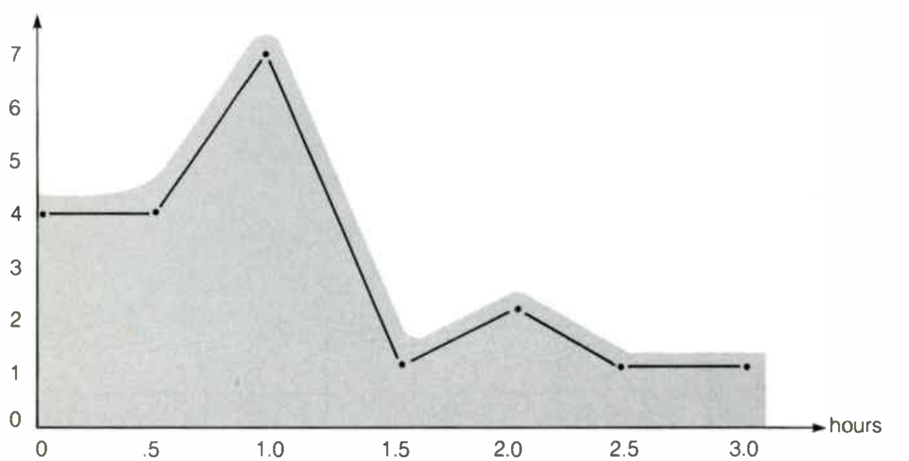
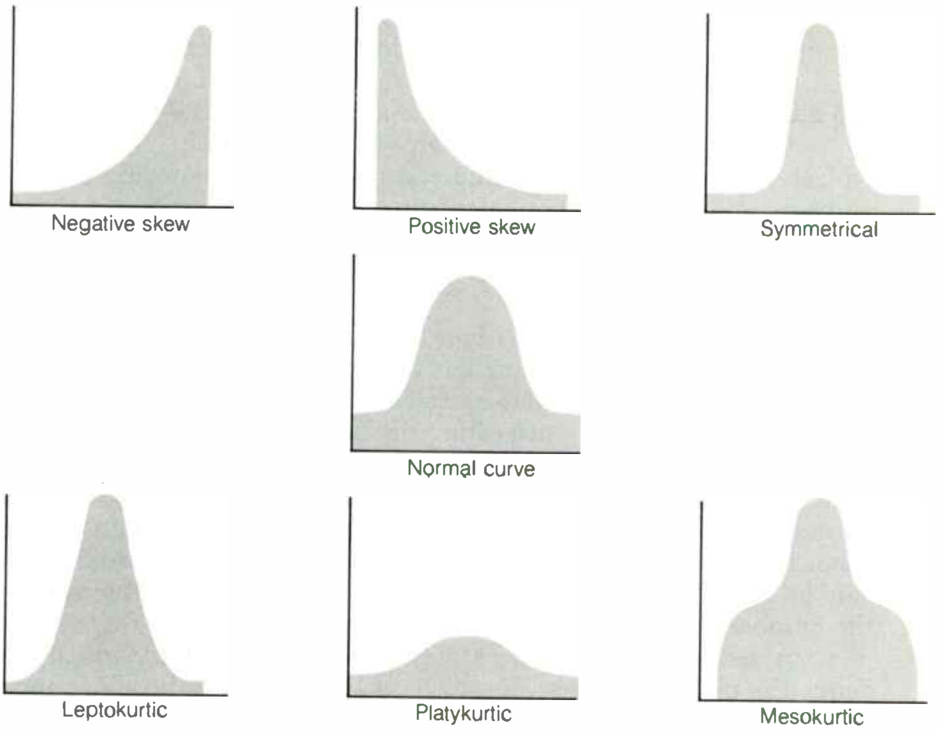


Figure 9.4 Skewness, kurtosis, and the normal curve



Skewness refers to the concentration of scores around a particular point on the x-axis. If this concentration lies toward the low end of the scale, with the tail of the curve trailing off to the right, the curve is *positively skewed* (or skewed to the right). If the tail of the curve trails off to the left, it is said to be *negatively skewed.* If the two halves of the curve are identical, then the curve is *symmetrical.*

Kurtosis refers to the peakedness of a frequency distribution, or the height of the curve along the y-axis. A *leptokurtic* distribution occurs when a large number of scores are clustered around the center of the distribution, while a *platykurtic* distribution is one where scores are spread over a wide area. A *mesokurtic* distribution is close to a normal distribution, but usually has two or more areas with a high concentration of scores.

A normal distribution is free from skewness and kurtosis. If researchers find that a frequency curve deviates excessively from the normal curve, they may have to manipulate or transform the data in some way in order to achieve a more normal distribution (transformation is discussed at the end of this chapter).

Summary Statistics

The data in Table 9.1 can be condensed still further through the use of **summary statistics.** These statistics help make data more manageable by measuring two basic tendencies of distributions: central tendency and dispersion.

Central tendency statistics answer the question, "What is a typical score?" They provide information about the grouping of the numbers in a distribution by calculating a single number that is characteristic of the entire distribution. Exactly what constitutes a "typical" score depends upon the level of measurement and the purpose for which the data will be used.

For every distribution, three types of characteristic numbers can be identified. One is the **mode** (*Mo*), or the score or scores occurring most frequently. There is no calculation necessary to determine the mode; it is found by inspecting the distribution. For the data in Table 9.1, the mode is 1.0. While easy to determine, the mode has some serious drawbacks as a descriptive statistic. It focuses attention on only one possible score and can thus camouflage important facts about the data when considered in isolation. This is illustrated by the data in Table 9.5: the mode is 70, but the most striking feature about the numbers is the way they cluster around 30.

A second characteristic score is the **median** (*Mdn*), which is defined as the point on a scale which 50% of the scores lie above and 50% below. To put it another way, the median is the midpoint of a distribution. If the distribution has an odd number of scores, the median is the middle score; if there is an even number, the median is a hypothetical score halfway between the two middle scores. To determine the median, one must order the scores from smallest to largest and then locate the midpoint by inspection. For example, here are nine scores:

0 2 2 5 6 17 18 19 67

Table 9.5 The Mode as a Potentially Misleading Statistic

Score	f
70	2
35–69	0
34	1
33	1
32	1
31	1
30	1
29	1
28	1
27	1
26	1

The median score is 6, since there are four scores above this number and four below it. Now consider these ten numbers:

$$\begin{array}{cccccccccc} 0 & 2 & 2 & 5 & 6 \quad\underset{11.5}{\uparrow}\quad 17 & 18 & 19 & 67 & 75 \end{array}$$

No score neatly bisects this distribution; to determine the median, the two middle scores must be added together and divided by two:

$$\text{Mdn} = \frac{6 + 17}{2} = 11.5$$

When there are many scores in the distribution that are the same, computing the median becomes a bit more complicated. Roscoe (1975) should be consulted for precise directions.

The third type of central tendency statistic is the **mean.** The mean is probably the most familiar summary statistic; it represents the average of a set of scores. Mathematically speaking, the mean is defined as the sum of all scores divided by *N*. Since the mean is widely used in both descriptive and inferential statistics, it will be described here in greater detail.

As a first step, some basic statistical notation is required:

X = any score in a series of scores

$\bar{X}$ (X-bar) = the mean

Σ (Greek capital letter sigma) = the sum

N = the total number of scores in a distribution

Using these symbols, the formula for the calculation of the mean is:

$$\bar{X} = \frac{\Sigma X}{N}$$

This equation indicates that the mean is the sum of all scores (ΣX) divided by the number of scores (N). Using the data in Table 9.1:

$$\bar{X} = \frac{20}{20}$$
$$= 1.00$$

If the data are contained in a frequency distribution, a slightly different formula is used to calculate the mean:

$$\bar{X} = \frac{\Sigma fX}{N}$$

In this case, X represents the midpoint of any given interval, and f is the frequency of that interval. Table 9.6 uses this formula to calculate the mean of the frequency distribution in Table 9.2.

Unlike the mode and the median, the mean takes into account all values in the distribution, making it especially sensitive to ex-

Table 9.6 Calculation of Mean from Frequency Distribution

Hours	f	fX
0.0	4	0.0
0.5	4	2.0
1.0	7	7.0
1.5	1	1.5
2.0	2	4.0
2.5	1	2.5
3.0	1	3.0
	N = 20	$\Sigma fX = 20$
	$\bar{X} = \frac{20}{20} = 1.00$	

treme scores or "outliers." It may be thought of as the score that would be assigned to each individual if the total were to be evenly distributed among all of them. It is also the only measure of central tendency that can be defined algebraically. As will be seen later, this allows the mean to be used in a wide range of situations. It also suggests that the data used to calculate the mean should be at the interval or ratio level (see Chapter 2).

In deciding which of the three measures of central tendency to report for a given set of data, the researcher must consider two main factors. First of all, the level of measurement used may determine the choice: if the data are at the nominal level, only the mode is meaningful; with ordinal data, either the mode or the median may be used. All three measures are appropriate for interval and ratio data, however, and a researcher may find it desirable to report more than one.

Second, the purpose of the statistic is important. If the ultimate goal is to describe a set of data, then the measure that is most typical of the distribution should be used. To illustrate, suppose the scores on a statistics exam were 100, 100, 100, 0, and 0. To say that the mean grade was 60 does not accurately portray the distribution; the mode or median would provide a more characteristic description.

The second type of descriptive statistics is used to measure **dispersion,** or variation. Measures of central tendency determine the typical score of a distribution; dispersion measures describe the way in which the scores are spread out about this central point. Dispersion measures can be particularly valuable when comparing different distributions. For example, suppose the average grades for two classes in research methods are exactly the same; however, one class has several excellent students and many poor students, while in the other class, all students are just about average. A measure of dispersion must be employed to reflect this difference. In many cases, an adequate description of a data set can be achieved by simply reporting a measure of central tendency (usually the mean) and an index of dispersion.

There are three measures of variation or dispersion; the simplest, **range** (R), is defined as the difference between the highest and lowest scores in a distribution of scores. The formula used to calculate the range is:

$$R = X_{hi} - X_{lo}$$

where X_{hi} = the highest score

X_{lo} = the lowest score

Since the range uses only two scores out of the entire distribution, it is not particularly descriptive of the data set. Additionally, the range often increases with sample size, since larger samples tend to include more extreme values. For these reasons, the range is seldom used in mass media research as the sole measure of dispersion.

A second measure, **variance,** provides a mathematical index of the degree to which scores deviate from, or are at variance with, the mean. A small variance indicates that most of the scores in the distribution lie fairly close to the mean; a large variance represents scores that are widely scattered. To put it another way, variance is directly proportional to the degree of dispersion.

To compute the variance of a distribution, the mean is subtracted from each score; these *deviation scores* are then squared, and the squares are summed and divided by N. The formula for variance (usually symbolized as S^2, although many textbooks use a different notation) is:

$$S^2 = \frac{\Sigma(X - \bar{X})^2}{N}$$

(In many texts, the expression $(X - \bar{X})^2$ is symbolized by x^2.) The numerator in this formula, $\Sigma(X - \bar{X})^2$, is called the *sum of squares.* Although this quantity is usually not reported as a descriptive statistic, the sum of squares is used in the calculation of several other statistics. An example using this variance formula is found in Table 9.7.

This equation may not be the most convenient formula for calculating variance, especially if N is large. A simpler, equivalent formula is:

$$S^2 = \frac{\Sigma X^2}{N} - \bar{X}^2$$

The expression ΣX^2 indicates that one should square each score and then sum the squared scores. (Note that this is not the same as $(\Sigma X)^2$, which means to sum all the scores and then square the sum.) An example of this formula is shown in Table 9.8, using the data from Table 9.7. Not surprisingly, the answers are the same.

Variance is a commonly used and highly valuable measure of dispersion. In fact, it is at the heart of one powerful technique, analysis of variance (see Chapter 11), which is widely used in inferential statistics. However, variance does have one minor inconvenience: it is

Table 9.7 Calculation of Variance

Score (X)	$\bar{X}$	$X - \bar{X}$	$(X - \bar{X})^2$
0.0	1.00	−1.00	1.00
0.0	1.00	−1.00	1.00
0.0	1.00	−1.00	1.00
0.0	1.00	−1.00	1.00
0.5	1.00	−0.5	.25
0.5	1.00	−0.5	.25
0.5	1.00	−0.5	.25
0.5	1.00	−0.5	.25
1.0	1.00	0	0
1.0	1.00	0	0
1.0	1.00	0	0
1.0	1.00	0	0
1.0	1.00	0	0
1.0	1.00	0	0
1.0	1.00	0	0
1.5	1.00	0.5	.25
2.0	1.00	1.00	1.00
2.0	1.00	1.00	1.00
2.5	1.00	1.50	2.25
3.0	1.00	2.00	4.00

$$S^2 = \frac{\Sigma(X - \bar{X})^2}{N} = \frac{13.5}{20} = 0.675$$

expressed in terms of squared deviations from the mean rather than in terms of the original measurements. In order to obtain a measure of dispersion that is calibrated in the same units as the original data, it is necessary to take the square root of the variance. This quantity, called the **standard deviation**, is the third type of dispersion measure. Usually symbolized as *S*, standard deviation is computed using either of the formulas shown below:

$$S = \sqrt{\frac{\Sigma(X - \bar{X})^2}{N}}$$

$$S = \sqrt{\frac{\Sigma X^2}{N} - \bar{X}^2}$$

Note that these two equations correspond to the respective variance formulas described above.

Standard deviation represents a given distance of the scores from the mean of a distribution. This figure can be especially useful in describing the results of standardized tests. For example, modern intelligence tests are constructed to yield a mean of 100 and a standard deviation of 15. A person with a score of 115 falls one standard deviation above the mean; a person with a score of 85 falls one standard deviation below the mean.

The notions of variance and standard deviation are easier to

Table 9.8 Calculation of Variance Using an Alternate Formula

Score (X)	X^2	
0.0	0	$S^2 = \frac{\Sigma X^2}{N} - \bar{X}^2$
0.0	0	$= \frac{33.50}{20} - (1.00)^2$
0.0	0	$= 1.675 - 1.00$
0.0	0	$= 0.675$
0.5	.25	
0.5	.25	
0.5	.25	
0.5	.25	
1.0	1.00	
1.0	1.00	
1.0	1.00	
1.0	1.00	
1.0	1.00	
1.0	1.00	
1.0	1.00	
1.5	2.25	
2.0	4.00	
2.0	4.00	
2.5	6.25	
3.0	9.00	
	$\Sigma X^2 = 33.50$	

understand if they are visualized. Figure 9.5 contains two pairs of frequency curves; the mean of each is denoted by a dotted line. Which curve in each pair would have the larger S^2 and S?

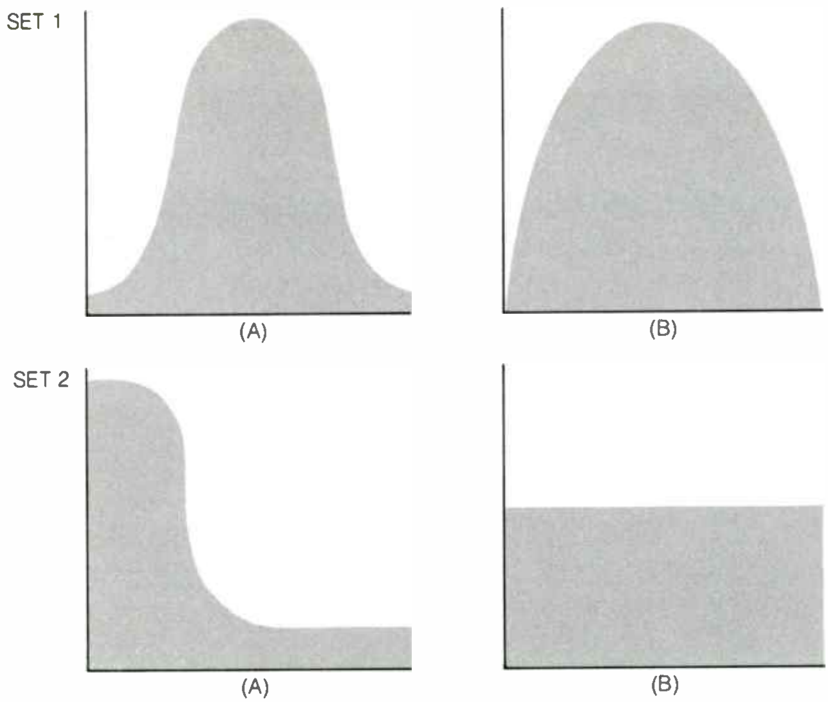
By determining the mean and the standard deviation, it is possible to derive **standard scores** (z) for any distribution of data. Standard scores are commonly used in psychology, education, and sociology, as well as in mass media research, to allow comparison of observations obtained by different methods.

A standard or "z" score indicates the placement of any score with respect to the mean, in terms of standard deviations above or below the mean. It can be calculated as follows:

$$z = \frac{X - \bar{X}}{S}$$

To illustrate, suppose that two roommates are in different sections of a research course. On a particular day each section is given a different exam. The first roommate scores a 40, and the second scores a 70. Surprisingly, the first roommate receives an "A," while the second receives a "C." In order to understand how the professor arrived at these marks, one must examine each student's standard score. It could be that the mean in the second roommate's class was 20 with a

Figure 9.5 Variance as seen in frequency curves



Which curve in each set represents the distribution with the larger S^2 and S?*

Answer: *Set 1 (B); Set 2 (A)

standard deviation of 5, while the mean in the first roommate's class was 72, also with a standard deviation of 5. Thus, the z score of Roommate 1 is 4.00, while Roommate 2's score is only -0.40. Clearly, with respect to the average grade in his or her section, Roommate 1 did extremely well. In fact, the z score indicates that he or she scored four standard deviations above average. On the other hand, Roommate 2's score was slightly below average; seen in this light, the "C" mark seems justified.

When any collection of raw scores is transformed into z scores, the resulting distribution possesses certain characteristics. Any score below the mean becomes a negative z score, while any score above the mean is positive. The mean of a distribution of z scores is 0, which is also the z score assigned to a person whose raw score equals the mean. The variance and standard deviation of z score distribution are both equal to 1.00. Standard scores are expressed in units of the standard deviation; thus, a z score of 3.00 means that the score is three standard deviation units above the mean.

Standard scores are used frequently in media research because

Table 9.9 The Mean and Standard Deviation of Raw and Standard Scores

Raw Score	z Score
2	−1.34
4	−.45
6	.45
8	1.34
$\bar{X} = 5$	$\bar{X} = 0$
$S = 2.24$	$S = 1.00$

they allow researchers to directly compare the performance of different subjects on tests using different measurements (assuming the distributions have similar shapes). They can literally be used to compare apples and oranges. Assume for a moment that the apple harvest for a certain year was 24 bushels per acre, compared to an average annual yield of 22 bushels per acre with a standard deviation of 10. During the same year, the orange crop yielded 18 bushels per acre, compared to an average of 16 bushels with a standard deviation of 8. Was it a better year for apples or oranges? The standard score formula reveals a z score of .20 for apples and .25 for oranges. Relatively speaking, oranges had a better year.

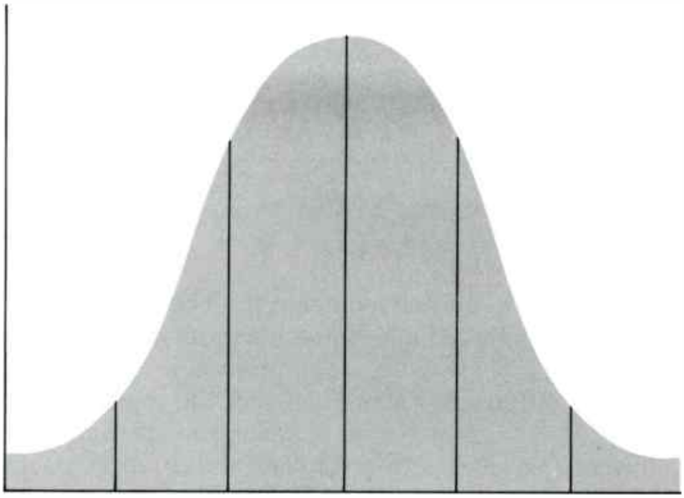
The Normal Curve

Standard scores not only enable comparisons to be made between dissimilar measurements, but, when used in connection with the normal curve, they also allow statements to be made regarding the frequency of occurrence of certain variables. Figure 9.6 shows an example of the familiar normal curve. As can be seen, the curve is symmetrical and achieves maximum height at its mean, which is also its median and mode. Also note that the curve in Figure 9.6 is calibrated in standard score units. When the curve is expressed in this way, it is called a *standard normal curve* and possesses all the properties of a z score distribution.

Statisticians have studied the normal curve closely in order to describe its properties. The most important of these is the fact that a fixed proportion of the area below the curve lies between the mean and any unit of standard deviation. The area under a certain segment of the curve is representative of the frequency of the scores that fall therein. Figure 9.7 portrays the areas contained under the normal curve between several key standard deviation units.

From the figure it can be determined that roughly 68% of the total area, and hence of the scores, lies within plus one and minus one standard deviations from the mean; 95% lies within plus two and minus two standard deviations; and so forth. Knowledge of this fact, together with the presence of a normal distribution, allows re-

Figure 9.6 The normal curve

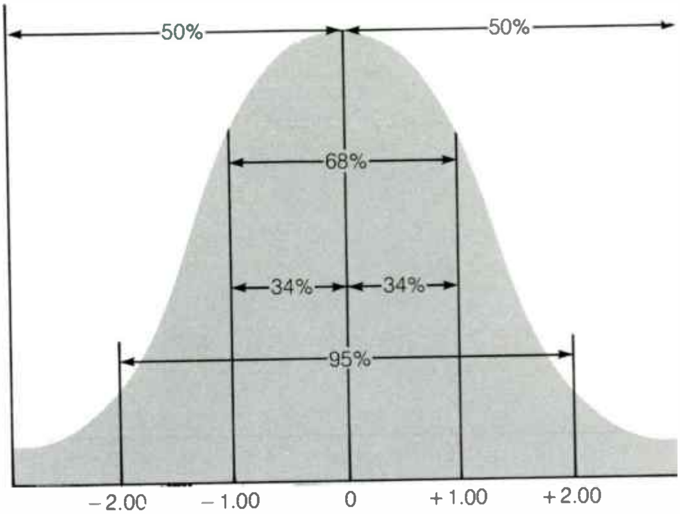


searchers to make useful predictive statements. For example, suppose that television viewing is normally distributed with a mean of 2 hours per day and a standard deviation of .5 hours. What proportion of the population watches between 2 and 2½ hours of TV? First, the raw scores are changed to standard scores: $\frac{2-2}{.5} = 0$ and $\frac{2.5-2}{.5} =$ 1.00. Figure 9.7 shows that approximately 34% of the area below the curve is contained between the mean and one standard deviation. Thus, 34% of the population watches between 2 and 2½ hours of television daily.

The same data can be used to find the proportion of the population that watches more than 3 hours of television per day. Again, the first step is to translate the raw figures into z scores. In this case, 3 hours corresponds to a z score of 2.00. A glance at Figure 9.7 shows that approximately 98% of the area under the curve falls below a score of 2.00 (50% in the left half of the curve plus about 48% from the mean to the 2.00 mark). Thus, only 2% of the population views more than 3 hours of television daily.

The normal curve is important because many of the variables that mass media researchers encounter are distributed in a normal manner, or normally enough that minor departures can be overlooked. Furthermore, the normal curve is an example of a probability distribution which becomes important in inferential statistics (see Chapter 11). Finally, many of the more advanced statistics discussed in later

Figure 9.7 Areas under the normal curve



chapters assume that the variable(s) under consideration is/are normally distributed.

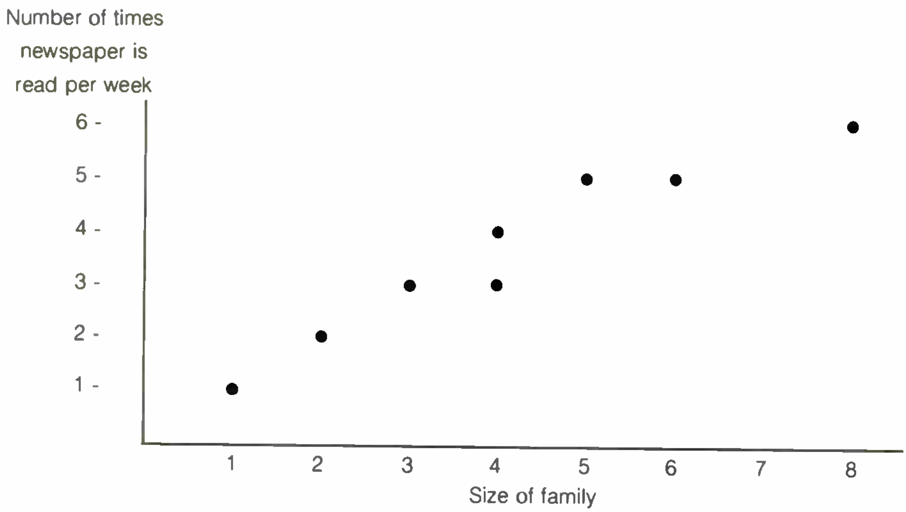
Basic Correlational Statistics

The discussion up to this point has dealt with analysis of a single variable. However, there are many situations where two different pieces of data must be examined simultaneously to determine if a relationship exists between them, that is, to discover whether the variables are covariant. For example, a researcher might hypothesize an association between the number of pictures on the front page of a newspaper and the total number of copies of the paper sold at newsstands. If his or her observations reveal that the more pictures there are, the more papers are sold, a relationship can be said to exist between the two variables. Numerical expressions of the degree to which two variables change in relation with one another are called *measures of association* or *correlation*.

When making two different measurements of the same person, it is common to designate one measure as the *x-variable* and the other as the *y-variable*. For example, in determining if there is a relationship between the size of a subject's family and the frequency with which that person reads a newspaper, the measure of family size could be the x-variable and the measure of newspaper reading the y-variable. Note that each subject in the group under study must be measured for both variables.

Figure 9.8 contains hypothetical data collected from a study of

Figure 9.8 Scattergram of family size and newspaper reading scores



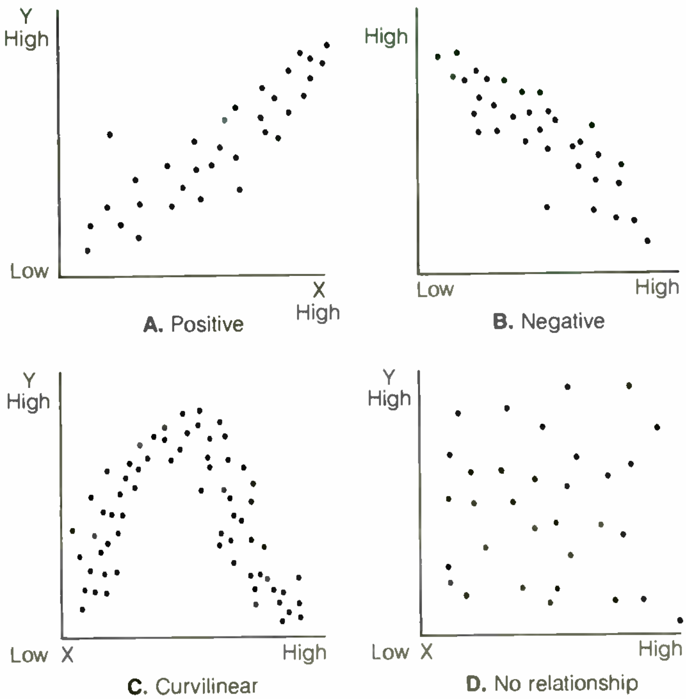
Subject	*Scores* (X) Family size	(Y) Number of times newspaper is read per week
A	1	1
B	2	2
C	3	3
D	4	3
E	5	4
F	6	5
G	7	5
H	8	6

eight subjects. The y-variable is the number of times per week the newspaper is read; the x-variable is the number of persons in the household. The scores are marked by a **scattergram,** a graphic technique for portraying a relationship between two or more variables. The two scores per subject are plotted on the scattergram. As indicated, family size and newspaper reading increase together. This is an example of a *positive relationship.*

An *inverse relationship* occurs when one variable increases while the other decreases. Sometimes the relationship between two variables is positive up to a point and then becomes inverse (or vice versa). When this happens, the relationship is said to be *curvilinear.* When there is no tendency for a high score on one variable to be associated with a high or a low score on another variable, the two are said to be *uncorrelated.* Figure 9.9 illustrates these relationships.

There are many statistics available to measure the degree of

Figure 9.9 Various scattergrams of possible relationships



relationship between two variables, but the most commonly used is the Pearson Product Moment Correlation, commonly symbolized as *r*. It varies between -1.00 and $+1.00$. A correlation coefficient of $+1.00$ indicates a perfect positive correlation: X and Y are completely covariant. A Pearson *r* of -1.00 indicates a perfect relationship in the negative direction. The lowest value that the Pearson *r* can achieve is 0.00. This represents absolutely no relationship between two variables. Thus, the Pearson *r* contains two pieces of information: (1) an estimate of the strength of the relationship, as indicated by the number; and (2) a statement about the direction of the relationship, as shown by the sign. Keep in mind that the strength of the relationship is solely dependent upon the number; strength of relationship must be interpreted in terms of absolute value. A correlation of $-.83$ is a stronger relationship than one of $+.23$.

The formula for calculating *r* looks foreboding; actually, however, it includes only one new expression:

$$r = \frac{N\Sigma XY - \Sigma X \Sigma Y}{\sqrt{[N\Sigma X^2 - (\Sigma X)^2][N\Sigma Y^2 - (\Sigma Y)^2]}}$$

X and *Y* stand for the original scores, *N* is the number of pairs of scores, and Σ again is the summation symbol. The only new term is

ΣXY, which stands for the sum of the products of each X and Y. To find this quantity, simply multiply each x-variable by its corresponding y-variable and add the results. Table 9.10 demonstrates a computation of r using the data from Figure 9.8. (The use of a calculator or computer is recommended when N is large, since the calculation of r can be tedious when many observations are involved.)

A correlation coefficient is a pure number—it is not expressed in feet, inches, or pounds, nor is it a proportion or percent. The Pearson r is independent of the size and units of measurement of the original data (in fact, the original scores do not have to be expressed in the same units). Because of its abstract nature, r must be interpreted with care. In particular, it is not as easy as it sounds to determine whether a correlation is large or small. Some writers have suggested various adjectives to describe certain ranges of r. For example, an r between .40 and .70 might be called a "moderate" or "substantial" relationship, while an r of .71 to .90 might be termed "very high." These labels are helpful, but may lead to confusion. The best advice is to consider the nature of the study. For example, an r of .70 between

Table 9.10 Calculation of r

Subject	X	X^2	Y	Y^2	XY
A	1	1	1	1	1
B	2	4	2	4	4
C	3	9	3	9	9
D	4	16	3	9	12
E	4	16	4	16	16
F	5	25	5	25	30
G	6	36	5	25	30
H	8	64	6	36	48
N = 8	$\Sigma X = 33$	$\Sigma X^2 = 171$	$\Sigma Y = 39$	$\Sigma Y^2 = 125$	$\Sigma XY = 145$

$(\Sigma X)^2 = 1{,}089$

$(\Sigma Y)^2 = 841$

$$r = \frac{(8)(145) - (33)(29)}{\sqrt{[(8)(171) - 1089][(8)(125) - 841]}}$$

$$= \frac{203}{\sqrt{(279)(159)}} = \frac{203}{(16.7)(12.6)}$$

$$= \frac{203}{210.42} = .965$$

r formula:

$$\frac{N\Sigma XY - \Sigma X\Sigma Y}{\sqrt{[N\Sigma X^2 - (\Sigma X)^2][N\Sigma Y^2 - (\Sigma Y)^2]}}$$

frequency of viewing television violence and frequency of arrest for violent crimes would be more than substantial; it would be phenomenal. Conversely, a correlation of .70 between two coders' timings of the length of news stories on the evening news is so low that the reliability of the study is questionable.

Additionally, correlation does not in itself imply causation. Newspaper reading and income might be strongly related, but this does not mean that earning a high salary "causes" people to read the newspaper. Correlation is just one factor in determining causality. Further, a large r does not necessarily mean that the two sets of correlated scores are equal. For example, the time spent reading a newspaper might have a correlation of .90 with the time spent viewing television news. This does not mean that people spend the same amount of time reading a newspaper as they do watching news. The Pearson r measures only covariation of the variables; it says nothing about magnitude.

Perhaps the best way to interpret r is in terms of the **coefficient of determination,** or the proportion of the total variation of one measure that can be determined by the other. This is calculated by squaring the Pearson r (r^2) to arrive at a ratio of the two variances: the denominator of this ratio is the total variance of one of the variables, while the numerator is that part of the total variance that can be attributed to the other variable. For example, if $r = .40$, then $r^2 = .16$: one variable explains 16% of the variation in the other. Or, to put it another way, 16% of the information necessary to make a perfect prediction from one variable to another is known. Obviously, if $r = 1.00$, then $r^2 = 100\%$; one variable allows perfect predictability of the other. The quantity $1 - r^2$ is usually called the **coefficient of nondetermination** because it represents that proportion of the variance left unaccounted for or unexplained.

Suppose that a correlation of .30 is found between a child's aggression and the amount of television violence the child views. This would mean that 9% of the total variance in aggression is accounted for by television violence. The other 91% of the variation is unexplained (except to the extent that it is not accounted for by the television variable). Note that the coefficient of determination is not measured on an equal interval scale: .80 is twice as large as .40, but this does not mean that an r of .80 represents twice as great a relationship between two variables as an r of .40. In fact, the r of .40 explains 16% of the variance, while the r of .80 explains 64%—four times as much.

The Pearson r can be computed between any two sets of scores. In order for the statistic to be a valid description of the relationship, however, several assumptions must be made: (1) that the data represent interval or ratio measurements; (2) that the relationship between X and Y is a linear one, not curvilinear; and (3) that the distributions of the x- and y-variables are symmetrical and comparable. (Pearson's r can also be used as an inferential statistic. When this is the case, it is necessary to assume that X and Y come from normally distributed populations with similar variances.) If these assumptions cannot validly be made, the researcher must use another kind of correlation coefficient, such as Spearman's rho or Kendall's W. For a thorough discussion of these and other correlation coefficients, the reader should consult Siegel (1956) or Nunnally (1978).

Inferential Statistics

As seen in Chapter 4, researchers are not generally content to obtain results that apply only to the sample studied. They are interested in results that can be generalized, or inferred, to the population from which the sample was drawn. The statistics that are used to determine the degree of generalizability are called **inferential statistics,** and are the most widely used of all mass media research techniques.

In order to discuss the basics of inferential statistics, it is necessary to introduce some new symbols to differentiate those statistics that indicate population parameters, or characteristics, from those used to describe a sample:

Characteristic	*Sample Statistic*	*Population Parameter*
Average	$\bar{X}$	μ (mu)
Variance	S^2	σ^2 (sigma squared)
Standard Deviation	S	σ (sigma)

It is also necessary to distinguish between three different types of distributions. A **sample distribution** is the distribution of some characteristic measured in the individuals or other units of analysis that were part of a sample. If a random sample of 1,500 college students were asked how many movies they attended in the last month, the resulting distribution of the variable "number of movies attended" would be a sample distribution, with a mean ($\bar{X}$) and variance (S^2). It is theoretically possible (though not practical) to ask the same question of every college student in the United States. This would create a **population distribution** with a mean (μ) and variance (σ^2). Ordinarily, the precise shape of the population distribution and the values of μ and σ^2 are unknown and are estimated from the sample. This estimate is called a **sampling distribution.**

In any sample drawn from a specified population, the mean of the sample, $\bar{X}$, will probably differ somewhat from the population mean, μ. For example, suppose that the average number of movies seen by all college students in the U. S. during the past month was exactly 3.8. It is unlikely that a random sample of ten students from this population would produce a mean of exactly 3.8. The amount that the sample mean differs from μ is called *sampling error* (see Chapter 4). If more random samples of ten were selected from this population, the values calculated for X that are close to the population mean would become more numerous than the values of X that are greatly different from μ. If this process were duplicated an infinite number of times and each mean placed on a frequency curve, the curve would form a sampling distribution.

Once the sampling distribution is identified, statements about the *probability* of occurrence of certain values are possible. There are many ways to define the concept of probability. Stated simply, the probability that an event will occur is equal to the relative frequency of occurrence of that event in the population under consideration (Roscoe 1975). To illustrate, suppose a large urn contains 1,000 table tennis balls, of which 700 are red and 300 white. The probability of drawing a red table tennis ball at random is 700/1,000, or 70%. It is also possible to calculate probability in situations where the relative frequency of occurrence of an event is determined theoretically. For example, what is the probability of randomly guessing the answer to a true/false question? One out of two, or 50%. What is the probability of guessing the right answer on a four-item multiple choice question? One out of four, or 25%. Probabilities can range from zero (no chance) to one (a sure thing). The sum of all the probable events in a population must equal 1.00, which is also the sum of the probabilities of an event occurring and of its not occurring. For instance, when a coin is tossed, the probability of its landing face up ("heads") is .50 and the probability of its not landing face up ("tails") is .50; .50 + .50 = 1.00.

There are two important rules of probability. The "addition rule" states that the probability that any one of a set of mutually exclusive events will occur is the sum of the probabilities of the separate events. (Two events are mutually exclusive if the occurrence of one precludes the other. For example, in the table tennis ball example, the color of the ball is either red or white; it cannot be both.) To illustrate the addition rule, consider a population where 20% of the people read zero magazines per month, 40% read only one, 20% read two, 10% read three, and 10% read four. What is the probability of selecting a person at random who reads at least two magazines per month? The answer is .40 (.20 + .10 + .10), the sum of the probabilities of the separate events.

The "multiplication rule" states that the probability of a combination of independent events occurring is the product of the separate probabilities of the events. (Two events are independent when the occurrence of one has no effect on the other. For example, getting "tails" on a flip of a coin has no impact on the next flip.) To illustrate the multiplication rule, calculate the probability that an unprepared student will correctly guess the right answers to the first four questions on a true/false test. The answer is the product of the probabilities of each event: .5 (chance of guessing right on Question 1) times .5 (chance of guessing right on Question 2) times .5 (chance of guessing right on Question 3) times .5 (chance of guessing right on Question 4) = .0625.

The notion of probability is important in inferential statistics because sampling distributions are a type of probability distribution.

Once the concept of probability is understood, a formal definition of "sampling distribution" is possible: A sampling distribution is *a probability distribution of all possible values of a statistic which would occur if all possible samples of a fixed size from a given population were taken*. For each outcome, the sampling distribution determines the probability of occurrence. For example, assume that a population consists of six college students. Their film viewing for the last month was as follows:

Student	*Number of films seen*
A	1
B	2
C	3
D	3
E	4
F	5

$\mu = (1 + 2 + 3 + 3 + 4 + 5)/6 = 3.00$

Suppose a study is made using a sample of two (N = 2) from this population. As is evident, there are only so many possible combinations that can be generated, assuming that sampling is done without replacement. Table 9.11 shows the possible outcomes.

The mean of this sampling distribution is equal to μ, the mean of the population. The likelihood of drawing a sample whose mean is 2.0 or 1.5 or any other value is found simply by reading the figure in the far right-hand column.

Table 9.11 is an example of a sampling distribution determined by empirical means. Many sampling distributions, however, are not derived by mathematical calculations but are determined theoretical-

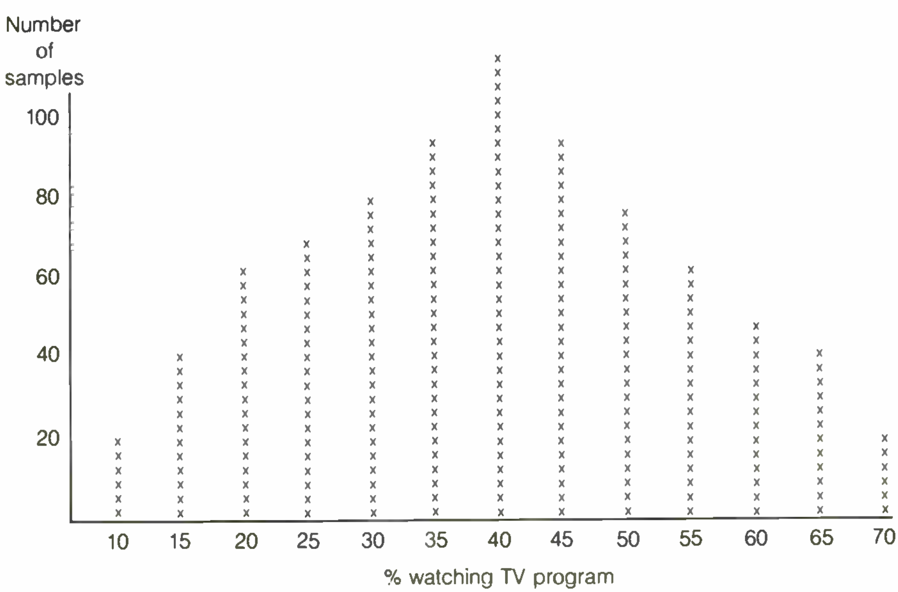
Table 9.11 Generating a Sampling Distribution Population = (1,2,3,3,4,5) N = 2

$\bar{X}$	# of possible sample combinations producing this $\bar{X}$	Probability of occurrence
1.5	2 (1,2) (2,1)	2/30 or .07
2.0	4 (1,3) (1,3) (3,1) (3,1)	4/30 or .13
2.5	6 (1,4) (2,3) (2,3) (3,2) (3,2) (4,1)	6/30 or .20
3.0	6 (1,5) (2,4) (3,3) (3,3) (4,2) (5,1)	6/30 or .20
3.5	6 (2,5) (3,4) (3,4) (4,3) (4,3) (5,2)	6/30 or .20
4.0	4 (3,5) (3,5) (5,3) (5,3)	4/30 or .13
4.5	2 (4,5) (5,4)	2/30 or .07
		1.00

Total number of possible sample combinations = 30 ly. For example, sampling distributions often take the form of a normal curve. When this is the case, the researcher can make use of all the information that is known about the properties of the normal curve. This can be illustrated by a hypothetical example using dichotomous data, or data with only two possible values. (This type of data is chosen because it makes the mathematics less complicated. The same logic applies to continuous data, but the computations are more elaborate.) Consider the case of a television rating firm attempting to estimate the total number of people in the population who saw a given program from the results of a sample. One sample of 100 people might produce an estimate of 40%, a second an estimate of 42%, and a third an estimate of 39%. If, after a large number of samples have been taken, the results are expressed as a sampling distribution, probability theory predicts that it would have the shape of the normal curve with a mean equal to μ. This distribution is shown in Figure 9.10.

In earlier discussions of the normal curve, the horizontal divisions along the base of the curve were expressed in terms of standard deviation units. With sampling distributions, this unit is called the *standard error* (*SE*) and serves as a criterion for determining the probable accuracy of an estimate. As is the case with the normal curve, roughly 68% of the sample will fall within plus one and minus one standard errors of the population mean, and 95% will fall within plus two and minus two standard errors.

Figure 9.10 Hypothetical sampling distribution



In most actual research studies, a sampling distribution is not generated by taking large numbers of samples and computing the probable outcome of each, and the standard error is not computed by taking the standard deviation of a sampling distribution of means. Instead, a researcher takes only one sample and uses it to estimate the population mean and the standard error. The process of inference from only one sample works in the following way: the sample mean is used as the best estimate of the population mean, and the standard error is then calculated from the sample data. Suppose that in the above example, it is found that 40 out of a sample of 100 people were watching a particular program. The mean, in this case symbolized as p since the data are dichotomous, is 40 percent (dichotomous data require this unique formula). The standard error in this dichotomous situation is calculated by the formula:

$$SE = \sqrt{\frac{pq}{N}}$$

where p = the proportion viewing

q = 1 − p

N = the number in the sample

In the example, the standard error is $\sqrt{\frac{(.6)(.4)}{100}}$ or .048. This means that about 68% of all possible samples will give estimates of 35.2%–44.8%, and 95% will give estimates of 30.4%–49.6%. There is only a .025 chance of getting a sample mean of less than 30.4%.

Instead of calculating the chances of a given sample mean falling two standard errors or more from the population mean, we could just as easily compute the chances of the population mean falling two or more standard errors from a given sample mean. In fact, the two procedures are simply mirror images of one another. It is just as probable that the sample mean will be two standard errors above the population mean as it is that the population mean will be two standard errors below the sample mean. This fact allows the researcher to construct *confidence intervals* (*CI*) from the results of one sample. There is a specified probability that the parameter will lie within a predetermined confidence interval. The most common confidence interval is the .95 level, which is expressed by the following formula:

$$.95CI = p \pm 1.96$$

where p = the proportion obtained in the sample

SE = the standard error

As an example, consider that a television ratings firm sampled 400 people and found that 20 percent of the sample were watching a certain program. What is the .95 confidence interval estimate for the population mean? The standard error is equal to the square root of

[(.20)(.80)]/400, or .02. Inserting this value into the above formula yields a confidence interval of .20 ± (2)(.02), or .16–.24. In other words, there is a .95 chance that the population average lies between 16% and 24%. There is also a 5% chance of error, that is, that μ lies outside this interval. If this 5% chance is too great a risk, it is possible to compute a .99 confidence interval estimate by substituting 2.58 for 2 in the formula (in the normal curve, 99% of all scores fall within ± 2.58 standard errors of the mean). For a discussion of confidence intervals using continuous data, the reader should consult Hays (1973).

The concept of sampling distribution is important to statistical inference. Confidence intervals represent only one way in which sampling distributions are used in inferential statistics. They are also an important feature of *hypothesis testing,* in which the probability of a specified sample result is determined under assumed population conditions (see Chapter 10).

Data Transformation

Most statistical procedures are based on the assumption that the data are normally distributed. Although many statistical procedures are "robust" or conservative, in their requirement of normally distributed data, there are instances where the results of studies using data that show a high degree of skewness or kurtosis may be invalid. The data used for any study should be checked for normality, a procedure accomplished very easily with most canned computer programs (see Chapter 16).

Most abnormal distributions are caused by outliers. When such cases arise, researchers can attempt to transform the data to achieve normality. Basically, transformation involves performing some type of mathematical adjustment to *each score* in order to try and bring the outliers closer to the group mean. This may take the form of multiplying or dividing each score by a certain number, or even taking the square root or log of the scores. It makes no difference what procedure is used (although some methods are more powerful than others), so long as the same method is employed for all the data.

There are a variety of transformation methods from which to choose, depending on the type of distribution found in the data. Rummel (1970) describes all of these procedures in great detail.

Summary

This chapter has introduced some of the more common descriptive and inferential statistics used by mass media researchers. Little attempt has been made to explain the mathematical derivations of the formulas and principles presented; rather, the emphasis here (as throughout the book) has been placed on understanding the reasoning behind these statistics and their applications. Unless researchers understand the logic underlying such concepts as "mean," "standard deviation," and "standard error," for example, the statistics themselves will be of little value.

Questions and Problems for Further Investigation

1. Find the mean, variance, and standard deviation for the following frequency distribution:

X	f
0	5
1	2
2	2
3	1
4	0

2. From a regular deck of playing cards, what is the probability of randomly drawing an ace? An ace *or* a nine? A spade *or* a face card?
3. Assume that scores on the Miller Analogies Test are normally distributed in the population with a μ of 50 and a population standard deviation of 5. What is the probability that:
   a. Someone picked at random will have a score between 50 and 55?
   b. Someone picked at random will score two standard deviations above the mean?
   c. Someone picked at random will have a score of 58 or higher?
4. Assume a population of scores consisting of the following: 2, 4, 5, 5, 7, and 9. Generate the sampling distribution of the mean if N = 2 (sampling without replacement).
5. Calculate r between the following set of scores:

X	Y
1	8
1	6
3	5
2	4
2	3
4	5
5	2
7	3

References and Suggested Readings

Blalock, H. M. 1972. *Social Statistics.* New York: McGraw-Hill.

Champion, D. J. 1981. *Basic Statistics for Social Research.* Boston: Houghton Mifflin.

Hays, W. L. 1973. *Statistics for the Social Sciences.* New York: Holt, Rinehart and Winston.

Nunnally, J. 1978. *Psychometric Theory.* New York: McGraw-Hill.

Roscoe, J. T. 1975. *Fundamental Research Statistics for the Behavioral Sciences.* New York: Holt, Rinehart and Winston.

Rummel, R. J. 1970. *Factor Analysis.* Chicago: Northwestern University Press.

Siegel, S. 1956. *Nonparametric Statistics for the Behavioral Sciences.* New York: McGraw-Hill.

Williams, F. 1979. *Reasoning with Statistics.* New York: Holt, Rinehart and Winston.

CHAPTER 10

Hypothesis Testing

It is a rare occurrence for a scientist to begin a research study without a problem or question to test. This is similar to holding a cross-country running race without telling the runners where to start. Both situations need an initial step: for the cross-country race, a starting line is required; the research study needs a statement or question to test. This chapter describes the procedures of developing research questions and the steps involved in testing them.

Research Questions and Hypotheses

Mass media research utilizes a variety of approaches to help answer questions. Some research is informal and seeks to solve relatively simple problems; some is based on theory and requires formally worded questions. All researchers, however, must start with some tentative generalization regarding a relationship between two or more of the variables under study. These generalizations may take two forms:

research questions and *statistical hypotheses*. The two are identical except for the aspect of prediction: hypotheses predict an experimental outcome; research questions do not.

Research Questions

Research questions are often used in problem- or policy-oriented studies where researchers are not specifically interested in testing the statistical significance of their findings. For instance, researchers analyzing television program preferences or newspaper circulation would probably be concerned only with discovering general indications and not with gathering data for statistical testing. However, research questions can be tested for statistical significance. They are not merely weak hypotheses; they are valuable tools for many types of research.

Research questions are frequently used in areas that have only been studied marginally, or not at all. Studies of this nature are classified as *exploratory research* because investigators have no idea as to what may be found. They do not have enough prior information to make predictions. Exploratory research is intended to search for data *indications* rather than to attempt to determine *causality* (Tukey 1962). The goal is to gather preliminary data in order to refine research questions and possibly to develop hypotheses.

Research questions may be stated as simple questions about the relationship between two or more variables, or about the components of a phenomenon. For example, researchers might ask: "What is the impact of spot advertising on the defections of voters from one party to another?" (Joslyn 1981); or: "What is the extent of employment of women on weekly newspapers in the United States, and the distribution of those women among different job types and levels?" (Holly 1979); or: "What percentages of sentences that appear in metropolitan daily newspaper stories about social issues fall into each of eight sentence types?" (Ryan 1979). These are broad questions which are not intended to predict an outcome.

There are, however, countless situations where researchers develop studies on the basis of existing theory and are thus able to attempt predictions about the outcome of the experiment. Doolittle (1979), for example, in his study of media use by older adults, hypothesized that: (1) news interest and use increases with age; and (2) educational attainment and income vary positively with age. Although these hypotheses may seem self-evident, Doolittle nevertheless had to consult previous studies before they could be generated. (Both of the hypotheses were supported by the study.)

In order to facilitate the discussion of research testing, the remainder of this chapter uses only the word hypothesis. But recall that excluding prediction, research questions and hypotheses are identical: each makes a tentative generalization concerning the relationship between two or more variables.

Purpose of Hypotheses

Hypotheses offer researchers a variety of benefits. First, they *provide direction* for a study. As indicated at the opening of the chapter, research begun without hypotheses offers no starting point; there is no indication of the sequence of steps to follow. Hypothesis development is usually the culmination of a rigorous literature review and emerges as a natural step in the research process. Without hypotheses, research would lack focus and clarity.

A second benefit of hypotheses is that they *eliminate trial and error research,* that is, research when investigators have no idea about the potential results and merely investigate a topic haphazardly in hopes of finding something significant. Hypothesis development requires researchers to isolate a specific area for study and eliminate any irrelevant topics. Trial and error research is time consuming and wasteful. The development of hypotheses eliminates this waste.

Hypotheses also *help rule out intervening and confounding variables.* Since hypotheses focus research to precise testable statements, other variables, whether relevant or not, are excluded. For instance, researchers interested in determining how the media are used to provide consumer information must develop a specific hypothesis stating what media are included, what products are being tested for what specific demographic groups, and so on. Through this process of narrowing, extraneous and intervening variables are eliminated or controlled. This does not mean that hypotheses eliminate all error in research; nothing can do that. Error in some form is always present in every study (see Chapter 2).

Finally, hypotheses *allow for quantification of variables.* As stated in Chapter 2, literally any concept or phenomenon is capable of quantification if put into an adequate operational definition. All terms used in hypotheses must have an operational definition. For example, in order to test the hypothesis: "There is a significant difference between recall of television commercials for subjects exposed to low frequency and high frequency broadcasts," researchers would need to operationally define "recall," "low frequency," and "high frequency." Words incapable of quantification cannot be included in a hypothesis.

In addition, some concepts can have a variety of definitions. One example of this is "violence." The complaint of many researchers is not that violence is incapable of quantification, but rather that it is capable of being operationally defined in more than one way. Therefore, prior to comparing the results of studies of media violence, it is necessary to consider the definition of "violence" used in each study. It is possible that contradictory results may be due to the definitions used and not the presence or absence of violence (see *Journal of Broadcasting* 1977, 21:3 for several articles on violence).

Criteria for Good Hypotheses

A useful hypothesis should possess at least four essential characteristics: (1) it should be compatible with current knowledge in the area; (2) it should follow logical consistency; (3) it should be in its most parsimonious form; and (4) it should be testable.

That hypotheses must be in harmony with current knowledge is obvious. If available literature strongly suggests one point of view, researchers who develop hypotheses that oppose this knowledge without basis only slow the development of knowledge in the area. For example, it has been demonstrated beyond a doubt that most people get their news information from television. It would be rather ludicrous for a researcher to develop a hypothesis suggesting that this is not true. There is simply too much evidence to the contrary.

The criterion of logical consistency means that if a hypothesis suggests that A = B and B = C, then A must also be equal to C. That is, if reading the *New York Times* implies a knowledge of current events, and a knowledge of current events means greater participation in social activities, then reading the *New York Times* should equate to a greater participation in social activities.

It should come as no surprise that hypotheses must be in their most parsimonious form. The concept of "the simpler, the better" (Occam's Razor) is stressed throughout this book. A hypothesis such as, "Intellectual and psychomotor creativity possessed by an individual positively coincides with the level of intelligence of the individual as indicated by standardized evaluative procedures measuring intelligence" is not exactly parsimonious. Stated simply, the hypothesis could read: "Psychomotor ability and IQ are positively related."

Developing an untestable hypothesis is clearly unproductive. Research is complicated enough without attempting to develop hypotheses which can add nothing to existing knowledge. For example, the hypothesis, "High school students with no exposure to television in their lifetime will develop paranoid tendencies after viewing television for 48 hours" could not be tested; where would researchers find a sample of high school students who have not seen television in their lifetime? Additionally, it is important to avoid value-laden concepts that are difficult or impossible to operationalize, such as: "Watching television is morally degrading."

The Null Hypothesis

Stated simply, the **null hypothesis** (also called the "alternative hypothesis" and the "hypothesis of no difference") asserts that the statistical differences or relationships being analyzed are due to chance or random error. The null hypothesis (H_0) is the logical alternative to the research hypothesis (H_1). For example, the hypothesis, "The level of attention paid to radio commercials is positively related to the amount of recall of the commercial" has its logical alternative, "The level of attention paid to radio commercials is not related to the amount of recall of the commercial."

In practice, researchers rarely state the null hypothesis. Since every research hypothesis does have its logical alternative, stating the null form is redundant (Williams 1979). However, the null hypothesis is always present and plays an important role in the rationale underlying hypothesis testing.

Testing Hypotheses for Statistical Significance

The procedures involved in hypothesis testing, or significance testing, revolve around either rejecting or accepting (supporting) the null hypothesis. That is, if H_0 is accepted, it is assumed that H_1 is rejected; and if H_0 is rejected, H_1 must be accepted.

In order to determine the statistical significance of a research study, the researcher must set a **probability level,** or significance level, against which the null hypothesis can be tested. If the results of the study indicate a probability lower than this level, the researcher can reject the null hypothesis. If the research outcome has a high probability, the researcher must support (or, more precisely, fail to reject) the null hypothesis. In reality, since the null hypothesis is not generally stated, acceptance and rejection apply to the research hypothesis and not the null hypothesis.

The probability level is expressed by a lowercase letter p (indicating probability), followed by a "less than" or "less than or equal to" sign, and then a value; for example: $p \leq .01$. This statement means that the null hypothesis is being tested at the .01 level of significance, and that the study will be considered statistically significant if the probability is equal to or lower than this level. Establishing a level of significance depends on the amount of error researchers are willing to accept (in addition to other factors peculiar to the particular research study). The question of error is discussed in greater detail later in the chapter.

It is common practice in behavioral research studies to set the probability level at either .01 or .05, which means that either one time or five times out of one hundred, the results of the study are based on random error or chance. There is no logical reason for using these figures; the practice has been followed for many years basically because Sir Ronald A. Fisher, the developer of the concept of significance testing, formulated tables based on the areas under the normal curve defined by these points. The current trend in many research areas is to set the significance level according to the purpose of the study rather than by general convention. Some studies may use .20 or .30, depending on the goals of the research. In exploratory research especially, more liberal levels are generally used; these are made more restrictive as further information is gathered.

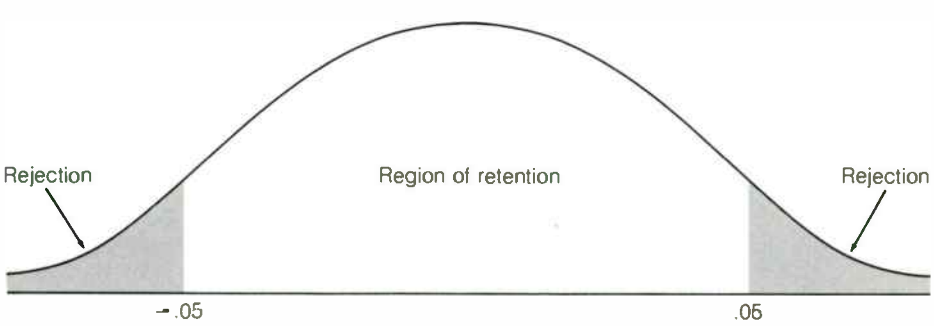
The **region of rejection** is that proportion of the area in a theoretical sampling distribution where the null hypothesis is rejected; its limits are defined by the probability level. The region of rejection is actually two regions, located at either end, or tail, of a sampling distribution, as shown in Figure 10.1.

The region between the two rejection values (or levels) is the area where the null hypothesis is retained. That is, if a research study calculates values outside the region of retention, the study is considered significant; if the value falls between the levels of rejection, the study is considered nonsignificant.

An important point in the region of rejection discussion refers to "one-tail testing" and "two-tail testing." These terms refer to the type of prediction made in a research study concerning the direction of the outcome. A one-tail test predicts that the results will fall in only one direction—either positive or negative. This approach is more stringent than the two-tail test which predicts results in both directions. Two-tail tests are generally used in situations where little information is available about the research area. One-tail tests are used in situations where researchers have more knowledge of the area and are able to more accurately predict the outcome of the study.

Consider, for example, a study of the IQ levels of a group of subjects who receive a special type of learning treatment, possibly a series of television programs on mathematics. The hypothesis is that the group, after viewing the programs, will have significantly different scores on a standardized math test than the remainder of the population, which has not seen the programs. The level of significance is set at .05, indicating that in order to reject the null hypothesis, the sample's mean test score must fall outside the boundaries in the normal distribution that are specified by the statement, "$p \leq .05$." These boundaries, or values, are determined by a simple computation. First, the critical values of the boundaries are found by consulting the normal distribution table (see Appendix). At the .05 level, these values are

Figure 10.1 The region of rejection



-1.96 and $+1.96$; the actual values can then be determined by means of the following formulas:

$$-1.96\ \alpha_m + \mu = \text{lower boundary}$$
$$+1.96\ \alpha_m + \mu = \text{upper boundary}$$

where α_m = the standard deviation of the distribution

μ = the population mean

We know from general knowledge that the population mean for IQ is 100 and the standard deviation is 15. Thus, the sample group must achieve a mean IQ score either lower than 70.60 or higher than 129.30 for the research study to be considered significant:

$$-1.96(15) + 100 = 70.60$$
$$+1.96(15) + 100 = 129.30$$

Using the normal distribution to demonstrate these boundaries, the area of rejection is illustrated in Figure 10.2.

Error

As with all steps in the research process, testing for statistical significance involves error. Two types of error are particularly relevant to hypothesis testing. These are known as **Type I error** and **Type II error;** Type I error is defined as rejecting a null hypothesis when it should be accepted, and Type II error as accepting a null hypothesis when it should be rejected. Another way to picture Type I and Type II error is shown in Figure 10.3.

The probability of making a Type I error is equal to the established level of significance and is therefore under the direct control of the researcher. That is, in order to reduce the probability of Type I

Figure 10.2 Regions of rejection for IQ problem

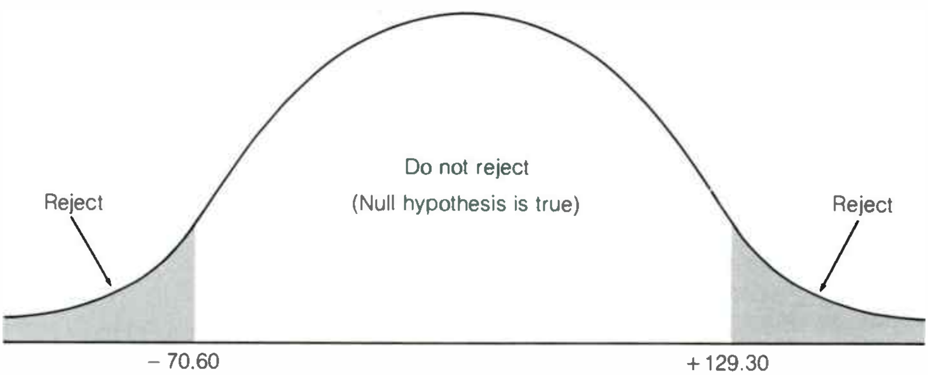


Figure 10.3 Type I and Type II error

	Reject H_0	Accept H_0
H_0 is true	Type I error	Correct
H_0 is false	Correct	Type II error

error, the researcher can simply set the level of significance closer to zero.

Type II error, signified by the symbol β, is a bit more difficult to conceptualize. The researcher does not have direct control over Type II error; instead, Type II error is controlled, although indirectly, by the design of the experiment. In addition, the level of Type II error is inversely proportional to the level of Type I error: as Type I error decreases, Type II error increases, and vice versa. The potential magnitude of Type II error depends in part upon the probability level and in part upon which of the possible alternative hypotheses actually is true. Figure 10.4 shows the inverse relationship between the two types of error.

As mentioned earlier, most research studies do not mention the null hypothesis, since it is generally assumed. There is a way to repre-

Figure 10.4 Relationship between Type I and Type II error

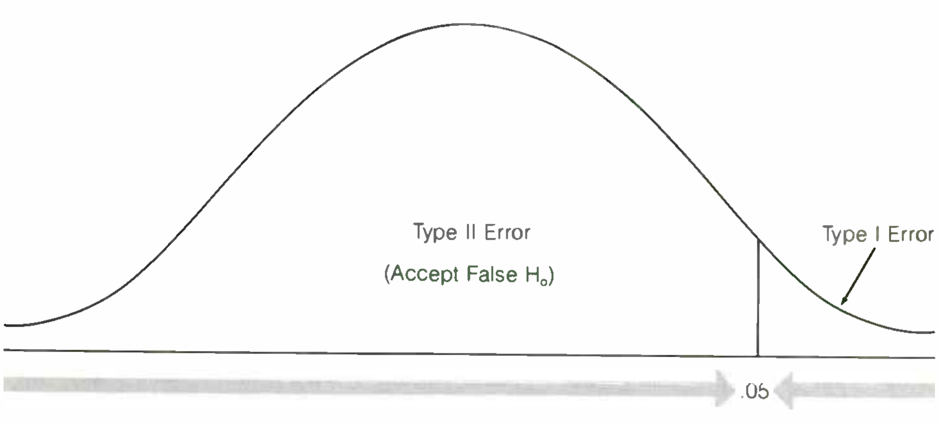


Figure 10.5 Type I and Type II error using the research hypothesis

	B	
A	Where one exists	Where none exists
Significant difference found	Correct	Type I error
No significant difference found	Type II error	Correct

sent Type I and Type II error in pictorial form without considering the null hypothesis. This approach may help to demonstrate the relationship between Type I and Type II error.

As Figure 10.5 demonstrates, the research hypothesis is used to describe Type I and Type II error instead of the null hypothesis. To read the table, start at the appropriate row on the left side, A, and then read the appropriate column under B; for example: "Significant difference found where none exists = Type I error."

One final way to explain Type I and Type II error is by using a hypothetical example. Consider a research study to determine the effects of a short-term public relations campaign for using safety belts in automobiles. Suppose that the campaign was actually highly successful and indeed changed the behavior of a majority of the subjects exposed to the campaign (this information is of course unknown to the researcher). If the researcher finds that a significant effect was created by the campaign, the conclusion is a correct one; if the researcher does not find a significant effect, he or she has committed a Type II error. On the other hand, if the campaign actually had no effect, but the researcher concludes that the campaign was successful, he or she has committed a Type I error.

The Importance of Significance

The concept of significance testing causes problems for many people. The main reason for this is that too many researchers overemphasize the importance of significance. When a study is conducted and the researcher finds that the results are nonsignificant, it is not uncommon for him or her to try and "talk around" the results—to deemphasize the fact that the results were not found to be significant. But there is really no need to follow this course of action.

There is no difference in the value of a study that finds significant results and a study which does not. Both studies provide valuable information. Discovering that some variables are not significant is just as important as determining what variables are significant. The nonsignificant study can save time for other researchers working in the same area by ruling out worthless variables. Nonsignificant research is important in collecting information about a theory of concept.

Also, there is nothing wrong with the idea of proposing a null hypothesis as the research hypothesis. For example, a researcher could formulate the following hypothesis: "There is no significant difference in comprehension of program content between a group of adults (age 18–49) with normal hearing that views a television program with closed-captioned phrases and a similar group that views the same program without captions." A scientific research study does not always have to test for significant relationships; it can also test for nonsignificance. However, sloppy research techniques and faulty measurement procedures can add to the error variance in a study and contribute to the failure to reject a hypothesis of no difference as well as jeopardize the entire study. This is a danger in using a null hypothesis as a substantive hypothesis.

Power

The concept of **power** is intimately related to Type I and Type II error. Power refers to the probability of rejecting the null hypothesis when an alternative is true. In other words, power indicates the probability that a statistical test of a null hypothesis will result in the conclusion that the phenomenon under study actually exists (Cohen 1969).

Statistical power is a function of three parameters: (1) probability level; (2) sample size; and (3) effects size. As we know, the probability level is under the direct control of the researcher and predetermines the probability of committing a Type I error. Sample size refers to the number of subjects utilized in an experiment. The most difficult concept is *effects size*. Basically, the effects size is the degree to which the null hypothesis is rejected; this can be stated either in general terms (such as *any* nonzero value) or in exact terms (such as .40). That is, when a null hypothesis is false, it is false to some degree; researchers can state generally that the null hypothesis is false and leave it at that, or they can specify exactly how false it is. The larger the effects size, the greater the degree to which the phenomenon under study is present (Cohen 1969). The problem is that researchers generally do not know the exact value of effects size. In this case, researchers can use one of three alternatives:

1. Estimate the effects size, based upon knowledge in the area of investigation or indications from previous studies in the area, or simply state the size as "small," "medium," or "large." (Cohen describes these values in greater detail.)

2. Assume an effects size of "medium."
3. Select a series of effects sizes and experiment.

Once the probability level, sample size, and effects size are known, researchers can consult power tables (located in most statistics books) to determine the level of power present in their study.

A determination of power is important for two reasons. First and most importantly, a low power level may prevent researchers from attaining statistical significance and thus cause a Type II error. This means that the otherwise nonsignificant results of a study may be made significant if the power of the statistical test is increased.

Second, a high power level may help in interpretation of research results. If an experiment just barely reaches the significance level but has high power, researchers can place more faith in the results. Without power figures, the researchers would have to be more hesitant in their interpretations.

Consideration of statistical power should be a step in all research studies. Although power is only an approximation, computation of the value helps control Type II error. In addition, as power increases, there is no direct effect on Type I error; power acts independently of Type I error. During the past few years researchers have begun to pay closer attention to statistical power.

Chase and Tucker (1975) conducted power analyses on articles published in nine communications journals. The authors found that 82% of the 46 articles analyzed had an average power for medium effects of less than .80 (the recommended minimum power value). In addition, over half of the articles had an average power of less than .50, which suggests a significant increase in the probability of Type II error.

Summary

Hypothesis development in scientific research is important because the process refines and focuses a research question. Rarely will a scientist begin a research study without some form of hypothesis. The purpose of this chapter was to introduce the process of developing hypotheses and to illustrate how they are tested and examined by mass media researchers. The following chapters describe various methods researchers use to test hypotheses and research questions.

Questions and Problems for Further Investigation

1. Develop three research questions and three hypotheses in any mass media area which could be investigated or tested.
2. What is your opinion about using more conservative levels of significance (.10 or greater) in exploratory research?

3. Conduct a brief review of published research in mass media. What percentage of the studies report the results of a power analysis calculation?
4. Explain the relationship between Type I and Type II error.
5. Under what circumstances might a researcher use a probability level of .001?
6. Assume that a researcher sets his/her *alpha* (significance) level at .02. The results of the experiment indicate that the null hypothesis cannot be rejected. What is the probability of a Type I error?

References and Suggested Readings

Chase, L. J., and Tucker, R. K. 1975. "A Power-Analytic Examination of Contemporary Communication Research." *Speech Monographs* 42:29–41.

Cohen, J. 1969. *Statistical Power Analysis for the Behavioral Sciences.* New York: Academic Press.

Doolittle, J. C. 1979. "News Media Use by Older Adults." *Journalism Quarterly* 56:2, 311–17.

Holly, S. 1979. "Women in Management of Weeklies." *Journalism Quarterly* 56:4, 810–15.

Joslyn, R. A. 1981. "The Impact of Campaign Spot Advertising on Voting Decisions." *Human Communication Research* 7:4, 347–60.

Roscoe, J. T. 1975. *Fundamental Research Statistics for the Behavioral Sciences.* New York: Holt, Rinehart and Winston.

Ryan, M. 1979. "Reports, Inferences and Judgments in News Coverage of Social Issues." *Journalism Quarterly* 56:3, 497–503.

Tukey, J. W. 1962. "The Future of Data Analysis." *Annals of Mathematical Statistics* 33:1–67.

Williams, F. 1979. *Reasoning with Statistics.* 2nd ed. New York: Holt, Rinehart and Winston.

CHAPTER 11

Inferential Statistics

Researchers often wish to do more than merely describe a sample; they want to infer their results to the population from which the sample has been taken. This chapter describes some of the basic inferential statistical methods used in mass media research and indicates ways in which these methods may help answer questions.

History of Small Sample Statistics

Samples were being employed in scientific research as long ago as 1627, when Sir Francis Bacon published an account of tests he had conducted measuring wheat seed growth in various forms of fertilizer. In 1763, Arthur Young began a series of experiments to discover the most profitable method of farming; and in 1849, James Johnston published a book called *Experimental Agriculture*, in which he provided advice on scientific research (Cochran 1976).

One of the most well-known investigators of the early twentieth century was William S. Gossett, who in 1908 attempted to quantify experimental results in a paper entitled, "The Probable Error of the Mean." Under the pen name "Student," Gossett published the results of small-sample investigations he had conducted while working in a Dublin brewery. The t-distribution statistics which Gossett developed were not widely accepted at the time; in fact, it was more than 15 years before other researchers began to take an interest in his work. The t-test, however, as will be seen, is now one of the most widely used statistical procedures in all areas of research.

R. A. Fisher provided a stepping stone from early work in statistics and sampling procedures to modern statistical inference techniques. It was Fisher who developed the concept of probability and established the use of the .01 and .05 levels of probability testing (see Chapter 10). Until Fisher, statistical methods were not generally perceived as practical in areas other than agriculture, where they were originally developed.

Nonparametric Statistics

Statistical methods are divided into two broad categories: **parametric** and **nonparametric.** Recall from Chapter 2 that a parameter is a population value such as the mean or variance. This definition describes the two broad categories of statistics: parametric statistics make assumptions about population parameters; nonparametric statistics make no such assumptions.

Another difference between the two categories is that nonparametric statistics are appropriate for nominal and ordinal data and parametric statistics are appropriate for interval and ratio data (see Chapter 2 for discussions of data forms). These two differences focus on one main point: parametric statistics are normally used to infer results from a sample to the population from which the sample was drawn; nonparametric statistics are generally not used for such an inference. However, the differences between nonparametric and parametric statistics have been questioned by many researchers during the past several years, and many researchers no longer consider that the two categories are distinctly different. Some statisticians and researchers argue that both methods can be used successfully with any type of data (see Roscoe 1975).

Chi-Square Goodness of Fit

There are many instances in mass media research where it is desirable to compare the *observed* frequencies of a phenomenon with those frequencies that might be *expected* or hypothesized. For example, a researcher might be attempting to determine whether the sales of television sets by four manufacturers in the current year are the same as sales for the previous year. He or she might advance the hypothesis, "Television set sales of four major manufacturers are significantly different this year from those of the previous year."

Suppose the previous year's television set sales were distributed as follows:

Type of Set	*Percent of Sales*
RCA	22%
Sony	36%
Quasar	19%
Zenith	23%

From these data, the investigator can calculate the expected frequencies (using a sample of 1,000) for each manufacturer's sales by multiplying the percentage of each sale by 1,000. The expected frequencies are:

Type of Set	*Expected Frequency*
RCA	220
Sony	360
Quasar	190
Zenith	230

Next, the researcher surveys a random sample of 1,000 households known to have purchased one of the four manufacturers' television sets during the current year. Assume the data from this survey indicate the following distribution:

Type of Set	*Expected Frequency*	*Observed Frequency*
RCA	220	180
Sony	360	330
Quasar	190	220
Zenith	230	270

The researcher now must interpret these data in such a way that he or she can state whether the change in frequency is actually *significant.* This can be done by reducing the data to a **chi-square statistic** and performing a test known as the chi-square "goodness of fit" test.

A chi-square (X^2) is simply a value showing the relationship between expected and observed frequencies. It is computed by means of the following formula:

$$X^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

where O_i = the observed frequencies

E_i = the expected frequencies

This means that the difference between each expected and observed frequency must be squared and then divided by the expected frequency. The sum of the quotients is the chi-square for those frequencies. For the above frequency distribution, chi-square is calculated as follows:

$$\chi^2 = \sum \frac{(O_1 - E_1)^2}{E_1}$$

$$= \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} + \frac{(O_3 - E_3)^2}{E_3} + \frac{(O_4 - E_4)^2}{E_4}$$

$$= \frac{(180 - 220)^2}{220} + \frac{(330 - 360)^2}{360} + \frac{(220 - 190)^2}{190} + \frac{(270 - 230)^2}{230}$$

$$= \frac{(-40)^2}{220} + \frac{(-30)^2}{360} + \frac{(30)^2}{190} + \frac{(40)^2}{230}$$

$$= \frac{1{,}600}{220} + \frac{900}{360} + \frac{900}{190} + \frac{1{,}600}{230}$$

$$= 7.27 + 2.50 + 4.73 + 6.95$$

$$= 21.45$$

Once the chi-square is known, the goodness of fit test can be performed to determine whether this value represents a significant difference in frequencies. To do this, the researcher must know two more values: the first is the probability level, which must be predetermined by the researcher; the second, called *degree of freedom (df)*, is the number of scores in any particular test which are free to vary in value. For example, if one has three unknown values (x, y, and z) such that $x + y + z = 10$, one would have two degrees of freedom: any two of the three variables may be assigned any value without affecting the total, but the value of the third will then be predetermined. Thus, if $x = 2$ and $y = 5$, z must equal 3. In the goodness of fit test, degrees of freedom are expressed in terms of $K - 1$, where K equals the number of categories. In the case of the television sales study, $K = 4$, and $df = 4 - 1 = 3$.

Next, the researcher must consult a chi-square table (see Appendix, Table 4). These tables are arranged by probability level and degrees of freedom. A portion of the chi-square table which is relevant to the hypothetical study is reproduced here to show how the table is used:

	Probability		
df	.10	.05	.01
1	2.706	3.841	6.635
2	4.605	5.991	9.210
3	6.251	7.815	11.345
4	7.779	9.488	13.277

If the calculated chi-square value equals or exceeds the value found in the table, the differences in frequency are considered statistically significant; if the calculated value is smaller, the results are considered to be nonsignificant.

In the television sales example, the researcher found a chi-square value of 21.45, with a degree of freedom of 3. Suppose he or she has established a probability level of .05; the chi-square table shows a value of 7.815 at this level when $df = 3$. Since 21.45 is greater than 7.815, the frequency difference is significant, and the hypothesis is accepted or supported: television set sales of the four manufacturers are significantly different in the current year from sales in the previous year.

The chi-square goodness of fit test can be used in a variety of ways to measure changes; for example, in studies of audience perceptions of advertising messages over a period of time, changes in television programming, or the results of public relations campaigns. Idsvoog and Hoyt (1977) used a chi-square test to analyze the professionalism and performance of television journalists. The authors were attempting to determine whether "professionalism" was related to several other characteristics, including the desire to look for employment, educational level, and job satisfaction. The results indicated that journalists classified as "high" professionals (based on a questionnaire) differed significantly from those journalists who were classified as either "medium" or "low" professionals.

There are limitations to the use of the goodness of fit test, however. Since this is a nonparametric statistical procedure, the variables must be measured at the nominal or ordinal level. The categories must be mutually exclusive, and each observation in each category must be independent from all others. Additionally, the chi-square distribution square is sharply skewed (see Chapter 9) for small samples, meaning that Type II error may occur: small samples may not produce significant results in cases where significant results could be obtained by using a larger sample. Most researchers suggest that each category contain at least five observations in order to avoid this problem.

As an alternative to the chi-square goodness of fit test, many researchers prefer the Kolomogorov-Smirnov test which is considered to be more powerful than the chi-square approach. The K-S procedure has many advantages over chi-square, most notably that a minimum number of expected frequencies in each cell is not required as in the chi-square test (see Winkler and Hays 1975 for more information about the Kolomogorov-Smirnov test).

Contingency Table Analysis

Another nonparametric statistical test that is often used in mass media research is the contingency table analysis, frequently called cross-tabulation or simply crosstabs. Crosstab analysis is basically an extension of the goodness of fit test, the primary difference being that two or more variables can be tested simultaneously. Consider a study to determine the relationship between a person's sex and his or her media usage habits in regard to obtaining information on new products. Suppose the researcher selects a random sample of 210 adults and obtains the following information:

Media Most Used for New Product Information

Sex		Radio	Newspapers	Television	
	Male	3	26	71	100
	Female	18	31	61	110
		21	57	132	

The next step is to calculate the expected frequencies for each cell. This procedure is similar to that used in the goodness of fit test, but it involves a slightly more detailed formula:

$$E_{ij} = \frac{R_i C_j}{N}$$

where E_{ij} = expected frequency for cell in row i, column j

R_i = sum of frequencies in row i

C_j = sum of frequencies in column j

N = sum of frequencies for all cells

Using this formula, the researcher in the hypothetical example can calculate the expected frequencies as follows:

$$\text{Male/Radio} = \frac{100 \times 21}{210} = \frac{2{,}100}{210} = 10$$

$$\text{Female/Radio} = \frac{110 \times 21}{210} = \frac{2{,}310}{210} = 11$$

and so forth. He or she then places each expected frequency in a small square in the upper right-hand corner of the appropriate cell.

Media Most Used for New Product Information

Sex		Radio	Newspapers	Television	
	Male	3 [10]	26 [27]	71 [63]	100
	Female	18 [11]	31 [30]	61 [69]	110
		21	57	132	

After the expected frequencies have been calculated, the investigator must compute the chi-square, using the following formula:

$$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Using the same example:

$$\chi_2 = \frac{(3 - 10)^2}{10} + \frac{(26 - 27)^2}{27} + \frac{(71 - 63)^2}{63} + \frac{(18 - 11)^2}{11} + \frac{(31 - 30)^2}{30} + \frac{(61 - 69)^2}{69}$$

$$= \frac{49}{10} + \frac{1}{27} + \frac{64}{63} + \frac{49}{11} + \frac{1}{30} + \frac{64}{69}$$

$$= 4.90 + .04 + 1.01 + .03 + .92$$

$$= 11.35$$

In order to determine statistical significance, the researcher must now consult the chi-square table. In a crosstab analysis, the degrees of freedom are expressed as $(R - 1)(C - 1)$, where R is the number of rows and C the number of columns. Assuming that $p \leq .05$, the chi-square value is listed as 5.991, which is lower than the calculated value of 11.35. Thus, there is a significant relationship between the sex of the respondent and the media he or she uses for new product information. The test indicates that the two variables are somehow related to one another, but it does not tell exactly how. In order to find this out, it is necessary to go back and examine the original crosstab data. Looking at the distribution, it is easy to see that females use radio more and television less than do males.

In the case of a 2 × 2 crosstab (where $df = 1$), computational effort is saved when the corresponding cells are represented by the letters A, B, C, and D, such as:

A	B
C	D

The following formula can then be used to compute the chi-square:

$$\chi^2 = \frac{N(AD - BC)^2}{(A + B)(C + D)(A + C)(B + D)}$$

Crosstab analysis has become a widely used statistical technique in mass media research, especially since the development of computer programs such as the *Statistical Package for the Social Sciences* (SPSS) (see Chapter 16). In addition to the chi-square, various other statistics can be used in crosstabs to determine whether or not the variables are statistically independent; alternatively, the data can be summarized by a variety of measures of association (see Nie et al. 1975).

Parametric Statistics

The statistical procedures discussed up to this point were nonparametric in nature; these methods are used primarily with nominal and ordinal data, and results are not generally intended to be generalized to the population. The following sections discuss parametric statistical methods which are intended for use with higher level data (interval and ratio) and results are intended to be generalized to the population from which the sample was drawn. The most basic parametric statistic is the t-test, a procedure widely used in all areas of mass media research.

The t-Test

In many research studies, two groups of subjects are tested: one group receives some type of treatment, and the other serves as the control. After the treatment is administered, both groups are tested, and the results are compared to determine if there is a statistically significant difference between the groups; that is, did the treatment have an effect on the results of the test? In cases such as this, the mean score for each group is compared through the use of a t-test, as described in Chapter 5.

The t-test is the most elementary method for comparing two groups' mean scores. A variety of t-test alternatives are available, depending on the problem under consideration and the situation of a particular research study. Variations of the t-test are available for testing independent groups, related groups, and situations where the population mean is either known or unknown (Champion 1981; Roscoe 1975).

The t-test assumes that the variables in the populations from which the samples are drawn are normally distributed (see Chapter 9). The test also assumes that the populations have homogeneity of variance; that is, that they deviate equally from the mean.

The basic formula for the t-test is relatively simple. The numerator of the formula is the difference between the sample mean and the hypothesized population mean, and the denominator is the estimate of the standard error of the mean (S_m):

$$t = \frac{\bar{X} - \mu}{S_m}$$

$$\text{where } S_m = \sqrt{\frac{SS}{(n-1)}}$$

$$SS = \Sigma(X - \bar{X})^2$$

One of the more popular forms of the t-test is the test for independent groups or means. This procedure is used in situations where two independent groups are studied for differences (the type of study described at the beginning of this section). The formula for the independent t-test is:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_{x_1 - x_2}}$$

where $\bar{X}_1$ = the mean for Group 1

$\bar{X}_2$ = the mean for Group 2

$S_{x_1 - x_2}$ = the standard error for the groups

The standard error is an important part of the t-test formula and is computed as follows:

$$S_{x_1 - x_2} = \sqrt{\left(\frac{SS_1 + SS_2}{n_1 + n_2 - 2}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

where SS_1 = the sum of squares for Group 1

SS_2 = the sum of squares for Group 2

N_1 = the sample size for Group 1

N_2 = the sample size for Group 2

To illustrate a t-test, imagine a research problem to determine the recall of two groups of subjects with regard to a television commercial for a new household cleaner. One group consists of 10 males and the other consists of 10 females. Each group views the commercial once and then completes a 15-item questionnaire concerning the information contained in the commercial. The hypothesis predicts a significant difference between the recall scores of males and females. The following data are collected:

Female Recall Scores

X	*x*	*x²(SS)*
4	−4	16
4	−4	16
5	−3	9
7	−1	1
7	−1	1
8	0	0
9	1	1
12	4	16
15	7	49
80		110

$\bar{X} = 8$

Male Recall Scores

X	*x*	*x²(SS)*
2	−4	16
3	−3	9
4	−2	4
4	−2	4
4	−2	4
6	0	0
6	0	0
10	4	16
13	7	49
60		106

$\bar{X} = 6$

Using the t-test formula, the next step is to compute the standard error for the groups by using the previous formula:

$$S_{x_1 - x_2} = \sqrt{\left(\frac{110 + 106}{10 + 10 - 2}\right)\left(\frac{1}{10} + \frac{1}{10}\right)}$$
$$= 1.55$$

The researcher then substitutes this standard error value in the t-test formula:

$$t = \frac{8 - 6}{1.55}$$
$$= 1.29$$

In order to determine whether the t value of 1.29 is statistically significant, a t-distribution table is consulted. The t-distribution is a family of curves closely resembling the normal curve. The portion of the t-distribution table relevant to the sample problem is reproduced below. Again, in order to interpret the table, two values are required: degrees of freedom and level of probability (a complete t-Distribution Table is included in the Appendix).

For purposes of the t-test, degrees of freedom are equal to $N_1 + N_2 - 2$, where N_1 and N_2 represent the sizes of the groups. In the example of advertising recall, df = 18 (10 + 10 − 2). If the problem is tested at the .05 level of significance, a t value of 2.101 is required for the study to be considered statistically significant. However, since the sample problem is what is known as a "two-tail test" (the hypothesis predicts only a difference between the two groups, not that one particular group will have the higher mean score), the required values are actually $t \leq -2.101$ *and* $t \geq 2.101$. The conclusion of the hypothetical problem is that there is no significant difference between the recall

t-Distribution Table

	Probability		
n	.10	.05	.01
1	6.314	12.706	63.657
2	2.920	4.303	9.925
–	–	–	–
–	–	–	–
–	–	–	–
17	1.740	2.110	2.898
18	1.734	2.101	2.878
19	1.729	2.093	2.861
–	–	–	–
–	–	–	–
–	–	–	–

scores of the female group and the recall scores of the male group, since t does not equal or exceed these values.

There are numerous examples of the t-test in mass media research that demonstrate the versatility of the method. For example, Genova and Greenberg (1979) employed this method to test the evidence that socioeconomic status and individual interests contribute significantly to the consumption of public affairs and news information. Jeffres (1978) used a t-test in his study of cable television and viewer selectivity to determine how motives for watching television change after a household is wired for cable.

Buerkel-Rothfuss and Mayes (1981) studied the differences between viewers and non-viewers of soap operas as regards their perceptions of various aspects of life (this is known as a cultivation analysis—see Chapter 8). The authors were interested in knowing whether soap opera viewers perceive life differently than non-viewers of soap operas. A sample of their results is shown in Table 11.1. For this test, the authors asked subjects to estimate the number of males and females in several different roles and positions in life. Subjects were asked: "Out of every 10 people, how many would you think are (*the role or position*)?"

From this table, it can be seen that there is a significant difference between average viewers and non-viewers regarding their perceptions of the number of males and females in particular societal roles and positions. The authors have also demonstrated that complicated research procedures are not necessary in order to uncover useful information.

Table 11.1 Buerkel-Rothfuss and Mayes Study of Viewers and Non-Viewers of Soap Operas (Mean scores and t values are listed)

Number of males estimated to be:	*Viewers (n = 206)*	*Non-viewers (n = 84)*	t
Doctors	2.2	1.3	6.5***
Lawyers	2.6	1.6	5.2**
Blue collar workers	5.3	5.0	1.3*
Divorced	4.6	3.9	3.0***
Happily married	4.8	4.8	−0.4*
Number of females estimated to be:			
Doctors	1.6	0.9	5.3***
Lawyers	1.8	1.1	4.7***
Housewives	6.3	5.5	3.3***
Divorced	4.6	4.0	2.8**
Happily married	4.9	4.8	0.2*

* Not significant

** Significant at $p < .01$

*** Significant at $p < .001$

Analysis of Variance

In many situations, researchers must investigate differences between more than two groups of subjects. In addition, there are also instances where researchers want to measure the effects of different degrees or levels of an independent variable (see Chapter 5). A t-test in these cases would not be adequate; what is required is an **analysis of variance** (ANOVA).

ANOVA is essentially an extension of the t-test. In fact, the two-sample ANOVA is mathematically equivalent to the t-test. The advantage of ANOVA, however, is that it can also be used in **factorial designs**, that is, research involving simultaneous analysis of two or more independent variables or factors. An ANOVA is classified according to the number of factors involved in the analysis: a *one-way* ANOVA investigates one independent variable, a *two-way* ANOVA investigates two independent variables, and so on. Also, several levels of any independent variable may be analyzed. Thus, a 2×2 ANOVA studies two independent variables, each with two levels.

ANOVA is a versatile statistic that is widely used in mass media research. The name of the statistic is somewhat misleading, however, because the most common form of ANOVA tests for significant differences between two or more group means, and has nothing to do with the analysis of differences of variance. Additionally, ANOVA breaks down the total variability of a set of data into its component *sources of variation;* that is, it "explains" the variation in a set of scores on one or more independent variables.

An ANOVA identifies or explains two types of variance: systematic and error. **Systematic variance** in data is that variance which is attributable to a known factor that systematically increases or decreases all the scores it influences. One such factor that is commonly identified in mass media research is sex: the fact that a subject is male or female will often systematically increase or decrease a given score. **Error variance** in data is created by an unknown factor that has most likely not been examined or controlled in the study. A primary goal of all research is to eliminate or control as much error variance as possible (a task that is generally easier to accomplish in the laboratory—see Chapter 5).

The ANOVA model assumes four things: (1) that each sample is normally distributed; (2) that variances for each group are equal; (3) that the subjects are randomly selected from the population; and (4) that the scores are statistically independent: they have no concomitant relationship with any other variable or score.

The ANOVA procedure begins with the selection of two or more random samples. Samples may be from the same or different populations. Each group is then subjected to different experimental treatments, followed by some type of test or measurement. The scores from the measurements are then used to calculate a ratio of variance, known as the F-ratio (*F*).

To understand this calculation, it is necessary to examine the procedure known as sum of squares (which was discussed briefly in Chapter 9) in greater detail. The sum of squares procedure involves squaring and summing either raw scores or deviation scores in order to eliminate the complicated problem of having to deal with negative numbers. The squaring process does not change the meaning of the data as long as the same procedure is used on all the data; it simply converts the data into a more easily interpreted set of scores.

In ANOVA, sums of squares are computed *between groups* (of subjects), *within groups* (of subjects), and in *total* (the sum of the between and within figures). The sums of squares between groups and within groups are divided by their respective degrees of freedom (as will be illustrated) to obtain a *mean square:* mean squares between (MS_b) and mean squares within (MS_w). The F-ratio is then calculated using the following formula:

$$F = \frac{MS_b}{MS_w}$$

where $MS_b\,df = K - 1$

$MS_w\,df = N - K$

$K =$ the number of groups

$N =$ the total sample

The F-ratio derived from the data is then compared to the value in the F-distribution table (found in the Appendix) that corresponds to the appropriate degrees of freedom and the desired probability level. If the calculated value equals or exceeds the tabled value, the ANOVA is considered to be statistically significant. The *F* table is similar to the *t* table and the chi-square table except that two different degrees of freedom are used, one for the numerator of the F-ratio and one for the denominator.

The ANOVA statistic can be demonstrated by using an example from advertising. Suppose that three groups of five subjects each are randomly selected to determine the credibility of a newspaper advertisement for a new laundry detergent. The groups are exposed to different versions of the advertisement that reflect varying degrees of design complexity: easy, medium, and difficult. The subjects are then asked to rate the advertisement on a scale of 1 to 10, with 10 indicating believable and 1 indicating not believable. The null hypothesis is advanced: "There is no significant difference in credibility among the three versions of the ad."

In order to test this hypothesis, the researchers must first calculate the three sums of squares: total, within, and between. The formulas for sums of squares (*SS*) are:

$$\text{Total}_{SS} = \Sigma X^2 - \frac{(\Sigma X)^2}{N}$$

$$\text{Within}_{SS} = \Sigma^2 - \frac{\Sigma\,(\Sigma X)^2}{N}$$

$$\text{Between}_{SS} = T_{SS} - W_{SS}$$

The scores for the three groups furnish the following data:

Group A (Easy)		*Group B (Medium)*		*Group C (Difficult)*	
X	X^2	X	X^2	X	X^2
1	1	4	16	6	36
2	4	5	25	7	49
4	16	6	36	7	49
4	16	6	36	8	64
5	25	8	64	10	100
16	62	29	177	38	298

$\Sigma X = 83\ (16 + 29 + 38)$

$\Sigma X^2 = 537\ (62 + 177 + 298)$

By inserting these figures in the formulas, the researchers are able to calculate the sums of squares as follows:

$$T_{SS} = \Sigma X^2 - \frac{(\Sigma X)^2}{N} = 537 - \frac{(83)^2}{15} = 537 - 459.2 = 77.8$$

$$W_{SS} = \Sigma X^2 - \frac{\Sigma(\Sigma X)^2}{N} = 537 - \frac{16^2}{5} - \frac{29^2}{5} - \frac{38^2}{5} = 537 - 508.2 = 28.8$$

$$B_{SS} = T_{SS} - W_{SS} = 77.8 - 28.8 = 49$$

With this information, the research team can deduce the mean squares between and within groups (SS/df), which can then be divided (MS_b/MS_w) to obtain the value of the F-ratio. These results are displayed in Figure 11.1.

Figure 11.1 Values for one-way ANOVA example

Sources of variation	df	Sums of squares	Mean square	F
Between groups	2 (K–1)	48.9	24.45	21.26
Within groups	12 (N–K)	13.8	1.15	xxxx
Total	14 (N–1)	62.7	xxxx	

Assuming a significance level of .05, the F-distribution table (see Appendix) for degrees of freedom of 2 and 12, indicates that the F-ratio must be equal to or greater than 3.89 in order to show statistical significance. It is clear that the calculated value of 10.2 is greater than 3.89; therefore, a significant difference in credibility among the three types of advertisements does exist, and the researchers must reject the null hypothesis.

Two-Way ANOVA

Researchers must often examine more than one independent variable in a study. For example, in the previous test, the researchers might have wished to investigate a second independent variable, product knowledge, at the same time. In this case, they would have to use a two-way ANOVA. The term "two-way" indicates the number of independent variables tested, in this case, two. A three-way ANOVA means that the study involves three independent variables, and so on. In addition, each independent variable can have several levels (easy, medium, hard) and this will generate a table with several rows and columns. In a two-way ANOVA, the researchers would gather the data and organize them in table form, as with the one-way ANOVA, but the two-way table would include both rows *and* columns, where each row and column represents an independent variable. The dependent variable score, represented by the letter X, for each subject is entered into each cell of the table. This procedure is demonstrated in Figure 11.2.

The two-way ANOVA can save time and resources, since two studies are actually being conducted simultaneously (one for each independent variable). In addition, it enables researchers to calculate two types of independent variable effects on the dependent variable: main effects and interactions (one-way ANOVA tests only for main effects). A **main effect** is simply the influence of an independent variable on the dependent variable. **Interaction** refers to the concomitant influence of two or more independent variables on the single dependent variable. For example, it may be found that a subject's educational background has no effect on media used for entertainment, but his or her education and socioeconomic status together may interact to create a significant effect.

The main effects plus interaction in a two-way ANOVA create a slightly different summary table than was shown for the one-way ANOVA. A two-way ANOVA summary table is shown in Figure 11.2.

Instead of computing only one F-ratio as in one-way ANOVA, a two-way ANOVA will compute four F-ratios, each of which is tested for statistical significance on the F-Distribution Table (see Appendix). The "Between Columns" (a main effect) represents the test of the independent variable levels located in the columns of a two-way ANOVA (from the previous example, this would be a test for the dif-

Figure 11.2 Two-way ANOVA table

	Group A (Easy)	Group B (Medium)	Group C (Hard)
No product knowledge	X_{111}, X_{112}	X_{121}, X_{122}	X_{131}, X_{132} . . . . .
High product knowledge	X_{211}, X_{212}	X_{221}, X_{222}	X_{311}, X_{321} . . . . .

The symbol "X" represents a dependent measurement score. The subscripts identify the subject who received that score. For example:

X_{111}
- Group A
- No Product Knowledge
- Subject Number 1

ferences between groups "easy," "medium," and "hard"). "Between Rows" is another main effects test which represents the significance between levels of the independent variable identified in the rows of the two-way ANOVA (product knowledge and no product knowledge). The "Interaction" section is the test for interaction between both independent variables in the study, and "Within Cells" tests for significant differences between each cell in the study to determine how each individual group performed in the analysis. F-ratios are not computed for the "Total" which accounts for the "x"s in the MS and F columns.

One final point regarding the two-way ANOVA bears discussion: in tests where a significant interaction between the independent variables is found, *only the interaction* is reported and interpreted. That is, regardless of whether the main effects of the individual independent variables are significant, they must be eliminated from the discussion if an interaction is present; interaction *always* takes precedence over main effects. This is because interaction is produced concomitantly by two or more independent variables, rendering the individual influence of each variable essentially meaningless. For example, consider a two-way ANOVA investigating the effects of room lighting and temperature on the ability to recall information from a television commercial. It might be that both of the independent variables are significant—that room temperature and room lighting do have some effect on recall. But if there is also an interaction—if temperature and lighting *together* have an effect on recall—then only this combined effect should be considered; any individual effects of the variables become irrelevant.

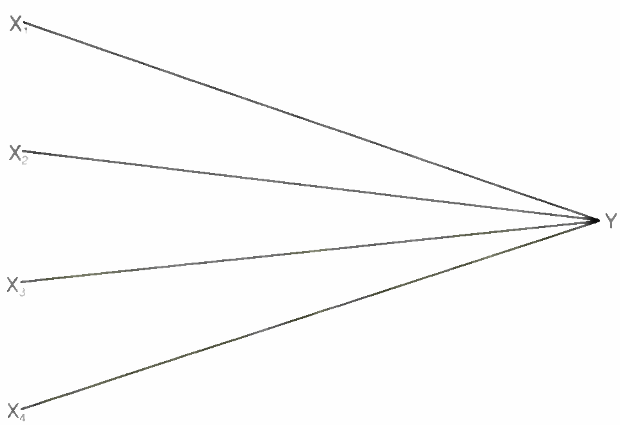
Multiple Regression

Multiple regression is another parametric technique used to analyze the relationship between two or more independent variables and a single dependent (criterion) variable. Although it is similar in some ways to an analysis of variance, the basic purpose of multiple regression is to *predict* the dependent variable, using information derived from an analysis of the independent variables.

In any research problem, the dependent variable is considered to be affected by a variety of independent variables. The primary goal of multiple regression is to develop a formula that accounts for, or explains, as much variance in the dependent variable as possible. It is widely used by researchers to predict success in college, sales levels, and so on. These dependent variables are predicted on the basis of *weighted linear combinations* of independent variables. A simple model of multiple regression is shown in Figure 11.3.

Linear combinations of variables play an important role in higher level statistics. In order to understand the concept of a weighted linear combination, consider the following methods of grading in two college classes. In one, the instructor determines each student's final grade by his or her performance on five exams: the scores on these exams are summed and averaged to obtain each final grade. A student might receive the following scores for the five exams: "B"

Figure 11.3 Multiple regression model



where X = an independent variable
Y = the predicted dependent variable

(3.0); "D+" (1.5); "B" (3.0); "B+" (3.5); and "A" (4.0); his or her final grade would then be a "B" (15/5 = 3.0). This grade is the dependent variable determined by the linear combination of five exam scores (the independent variables). No test is considered more important than another; hence, the linear combination is not said to be weighted (except in the sense that all of the scores are "weighted" equally).

In the second class, the instructor also determines the final grades by students' performances on five exams; however, he or she specifies that the first exam is to count 30%, the last exam 40%, and the remaining three exams 10% each in the determination. A student with the same five scores as above would thus receive a final grade of 3.3. Again, the scores represent a linear combination, but it is a weighted linear combination: the first and last exam contribute more to the final grade than do the other tests.

This latter "grading system" is used in multiple regression: the independent variables are weighted and summed in order to predict a dependent variable. The weight of each variable in a linear combination is referred to as its **beta weight.**

A multiple regression formula may involve any number of independent variables, depending on the complexity of the dependent variable. A simple formula of this type might look as follows (hypothetical values are used):

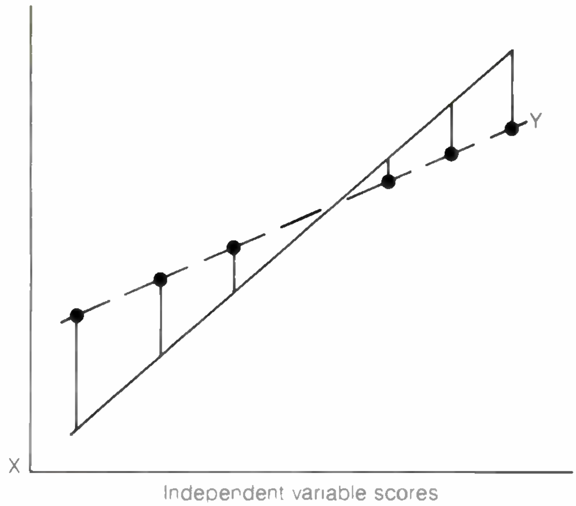
$$\hat{Y} = .89X_1 + 2.5X_2 - 3$$

where $\hat{Y}$ = the predicted score or variable
X_1 = Independent variable 1
X_2 = Independent variable 2

The number 3 in the formula is a constant subtracted from each subject's scores, and is derived as part of the multiple regression formula. All formulas produced by multiple regression analyses represent a line in space; that is, the dependent variable is interpreted as a linear combination, or line, of independent variables. The slope of this line is determined by the beta weights (also known as *regression coefficients*) assigned to the variables (see Thorndike 1978; Cohen and Cohen 1975). The goal of the researcher is to derive a formula for a line that coincides as nearly as possible with the *true* line (a mathematically determined line that represents a perfect prediction) of the dependent variable: the closer the computed line comes to the true line, the more accurate the prediction will be.

The basic computational procedure involved in multiple regression to predict a line is known as the principle of least squares. This procedure, shown in Figure 11.4, bears discussion. The multiple regression problem first begins by gathering data for the independent variables. The values for these variables are placed (by the computer) on a graph (represented by dots in Figure 11.4). The computer will then determine the line equation for the plotted data points such that the line passes through, or near, the greatest number of points. This

Figure 11.4 Principle of least squares



computed line is then compared to the true, or perfect, line to determine the accuracy of the predicted line developed from the data. The closer the computed line is to the true line, the more accurate the prediction will be.

The solid line in Figure 11.4 represents the true value; the broken line is plotted according to the multiple regression equation. As can be seen in 11.4, the independent variables (represented by dots) do not fall directly on the solid line, but rather are located at various distances from it. The principle of least squares requires that these distances be measured and the resulting values squared (to eliminate negative values) and then summed, as the two lines approach coincidence. Thus, the principle provides a measure of the predictive value of a multiple regression formula: the smaller the sum of squares, the higher the accuracy with which the formula predicts the dependent variable. As Kerlinger and Pedhazur (1973) noted:

> In any prediction of one variable from other variables there are errors of prediction. All scientific data are fallible. The data of the behavioral sciences are considerably more fallible than most data of the natural sciences. This really means that errors of prediction are larger and more conspicuous in analysis. In a word, error variance is larger. The principle of least squares tells us, in effect, to so analyze the data that the squared errors of prediction are minimized.

Another important value that must be calculated in a multiple regression analysis is the *coefficient of correlation* (R), which represents the product-moment correlation (see Chapter 9) between the

Figure 11.5 Drew and Reeves multiple regression table

Predictor Variables	Beta Weights
Like program	.15**
Credibility	.10*
Informational Content	.39***
Like Story	.25***
Multiple R	.546
R^2	.298

*$p < .05$
**$p < .01$
***$p < .001$

predicted $\hat{Y}$ score and the weighted linear combination of the X scores. The square of this coefficient (R^2) indicates the proportion of variance in the dependent variable that is accounted for by the predictor variables. The higher the R^2 (that is, the closer the figure is to 1.00), the more accurate the prediction is considered to be.

Drew and Reeves (1980) conducted a multiple regression analysis to determine what factors affect the way children learn from television news stories. They defined the dependent variable, "learning," in terms of performance on a 10-point questionnaire regarding a news program the children watched in an experimental setting. The selection of independent variables was based on the results of previous studies; the four that they decided to measure were: (1) whether the children liked the program; (2) whether the children liked the particular news story; (3) the credibility of the program; and (4) the informational content of the particular story.

The results, shown in Figure 11.5, indicate that all of the independent variables were statistically significant in their relation to learning. As the beta weights show, "informational content" seems to be the best predictor of learning, while "credibility" accounts for the least amount of variance. The multiple *R* of .546 could be considered highly significant; however, since it means that only 29% ($.546^2$) of the variance in the dependent variable was accounted for by the four predictor variables, this value may not substantially explain the variance.

Partial Correlation

Partial correlation is a method researchers use when they believe that a confounding or spurious variable may affect the relationship between the independent variables and the dependent variable: if such an influence is perceived, they can "partial out" or control the confounding variable. For example, consider a study of the relationship between exposure to television commercials and the purchase of the advertised products. The researchers decide to show two different commercials for a liquid laundry detergent (a "straight" version, with no special video or audio effects, and a "hard sell" version that does use special effects) to two groups of subjects (people who use only powdered detergent and people who use only liquid detergent). The study design would look as shown in Figure 11.6.

Assume the results show a very low correlation, indicating that any prediction made on the basis of these two variables would be very tenuous. The researchers should consider whether a confounding variable might be responsible. An examination might reveal, for example, that the technicians had problems adjusting the color definition of the recording equipment; instead of its natural blue color, the detergent appeared dingy brown on the television screen. The study could be repeated to control (statistically eliminated) for this variable by filming new commercials with the color controls properly adjusted. The design for the new study would be as shown in Figure 11.7.

The partial correlation statistical procedure would enable the researchers to determine the influence of the controlled variable. It might be that the correlation would increase from the original study using the new statistical method.

Cutler and Danowski (1981) used partial correlation in their study of older persons' use of television. The authors found it necessary, on the basis of suggestions from previous analyses, to control for sex and education when determining the correlation between political interest and television use. When these variables were partialed out (controlled), they found that media use varied with the subject's age and when the media were used during the campaign.

Figure 11.6 Product purchase study

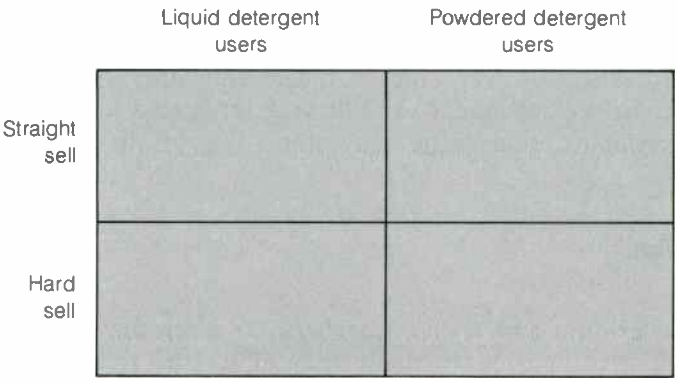


Figure 11.7 Product purchase study: partial correlation analysis

	Brown soap ads		Blue soap ads	
	Liquid detergent users	Powdered detergent users	Liquid detergent users	Powdered detergent users
Straight sell				
Hard sell				

Summary

Mass media research has made great strides during the past several years, in terms of both the number of research studies completed and the types of statistical methods used. The purpose of this chapter has been to introduce some of the more widely used inferential statistical procedures involving one dependent variable and one or more independent variables. The information is intended to help beginning researchers in reading and analyzing published research.

The emphasis in this chapter is on *using* statistical methods rather than on the statistics themselves. The basic formula for each statistic is briefly outlined so that beginning researchers can understand how the data are derived; the goal, however, has been to convey a knowledge of how and when to use each procedure. It is important that researchers be able to determine not only what the problem or research question is, but also which statistical method most accurately fits the requirements of a particular research study.

Questions and Problems for Further Investigation

1. Design a mass media study for which a chi-square analysis is appropriate.
2. In the chi-square example of television set sales, assume that the observed sales frequencies are 210 (RCA), 350 (Sony), 200 (Quasar), and 240 (Zenith). What is the chi-square value? Is it significant?
3. What are the advantages of using an ANOVA over conducting several separate t-tests of the same phenomena?
4. How could multiple regression be used to predict a subject's television viewing, radio listening and newspaper reading behavior?

5. The t-test is frequently used in natural science investigations. In the early years of mass media research, the t-test was also considered a staple research tool. During the past several years, however, this practice has changed: researchers no longer seem to rely on the t-test. Why do you think this is true?

References and Suggested Readings

Atwood, L. E., and Sanders, K. R. 1976. "Information Sources and Voting in a Primary and General Election." *Journal of Broadcasting* 20:291–301.

Buerkel-Rothfuss, N. L., and Mayes, S. 1981. "Soap Opera Viewing: The Cultivation Effect." *Journal of Communication* 31:108–15.

Champion, D. J. 1981. *Basic Statistics for Social Research.* New York: Macmillan.

Cochran, W. G. 1976. "Early Development of Techniques in Comparative Experimentation." *On the History of Statistics and Probability,* ed. by D. B. Owen. New York: Marcel Dekker.

Cohen, J., and Cohen, P. 1975. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences.* Hillsdale, N.J.: Lawrence Erlbaum Associates.

Cutler, N. E., and Danowski, J. A. 1981. "Process Gratification in Aging Cohorts." *Journalism Quarterly* 57:269–76.

Drew, D., and Reeves, B. 1980. "Learning from a Television News Story." *Communication Research* 7:121–35.

Genova, B. K. L., and Greenberg, B. S. 1979. "Interests in News and the Knowledge Gap." *Public Opinion Quarterly* 43:79–91.

Idsvoog, K. A., and Hoyt, J. L. 1977. "Professionalism and Performance of Television Journalists." *Journal of Broadcasting* 21:97–109.

Jeffres, L. 1978. "Cable TV and Viewer Selectivity." *Journal of Broadcasting* 22:167–78.

Kerlinger, F. N., and Pedhazur, E. J. 1973. *Multiple Regression in Behavioral Research.* New York: Holt, Rinehart and Winston.

Krull, R., and Husson, W. 1980. "Children's Anticipatory Attention to the TV Screen." *Journal of Broadcasting* 24:35–48.

Metallinos, N., and Tiemens, R. 1977. "Asymmetry of the Screen: The Effect of Left versus Right Placement of Television Images." *Journal of Broadcasting* 21:21–34.

Nie, N. H., *et al.* 1975. *Statistical Package for the Social Sciences.* New York: McGraw-Hill.

Presser, S., and Schuman, H. 1980. "The Measurement of a Middle Position in Attitude Surveys." *Public Opinion Quarterly* 44:70–85.

Reeves, B., and Miller, M. 1977. "A Multidimensional Measure of Children's Identification with Television Characters." *Journal of Broadcasting* 22:71–86.

Roscoe, J. T. 1975. *Fundamental Research Statistics for the Behavioral Sciences.* New York: Holt, Rinehart and Winston.

Thorndike, R. M. 1978. *Correlation Procedures for Research.* New York: Gardner Press.

Wakshlag, J., and Greenberg, B. S. 1979. "Programming Strategies and the Popularity of Television Programs for Children." *Human Communication Research* 6:58–68.

Winer, B. J. 1971. *Statistical Principles in Experimental Design.* New York: McGraw-Hill.

Winkler, R., and Hays, W. 1975. *Statistics: Probability, Inference, and Decision.* New York: Holt, Rinehart and Winston.

CHAPTER 12

Multivariate Statistics

The discussions of statistics in the previous three chapters have dealt with *univariate* procedures, which are used to investigate the relationship between one or more independent variables and a single dependent variable. This chapter is a preliminary discussion of **multivariate statistics**, methods that allow analysis of several independent variables *and* several dependent variables in a single study.

The rationale for using multivariate statistics in mass media research is quite simple. Both human behavior and the media are complex systems of interacting variables. There are probably few situations, if any, where one dependent variable accurately represents a phenomenon or is solely responsible for a particular attitude or behavior. Instead, a series of dependent variables, correlated to some degree, act together to produce or represent a phenomenon, or create an attitude or behavior.

Because researchers accept the idea of interacting variables affecting interrelated phenomena, many of them select multivariate statistical methods to analyze data rather than the more traditional, and limited, univariate approaches. This does not mean that they consider univariate procedures invalid, but rather that multivariate statistics are generally more useful in media research.

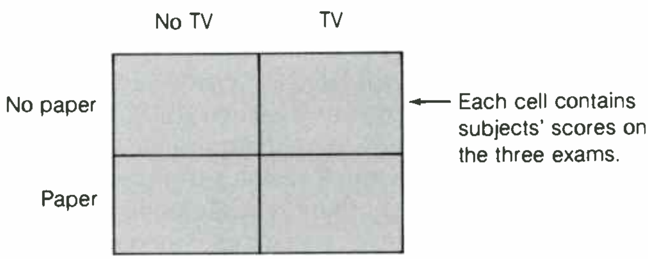
The present reliance upon multivariate analysis in research is based on several insights: (1) any given experimental manipulation affects several different, but partially related, areas of an individual's behavior, and univariate analysis is capable of investigating only one of these relationships at a time (Harris 1975); (2) because human beings are multidimensional, it seems reasonable to study them on several dimensions simultaneously, rather than to focus on a single, often arbitrarily chosen variable (Tucker 1982); (3) measurements taken on the same individual are correlated by virtue of their common origin and thus lend themselves to simultaneous study (Tucker 1982); (4) multivariate statistical methods are parsimonious and save time, money, and resources, since it is much simpler to investigate several correlated variables simultaneously than to study them one at a time; and (5) multivariate analysis allows researchers to investigate variables as *structures* or *constructs* rather than as the individual components of a structure or construct (Cattell 1966).

Multivariate statistics, however, do have some disadvantages. First of all, they are difficult to use compared to univariate methods. It is not possible to sit down and learn the procedures involved in factor analysis, for example, as one can do with a t-test; multivariate methods require extensive reading and trial-and-error work with a computer. Additionally, the interpretation of multivariate results is equally difficult in many instances. The results of a t-test are relatively straightforward; two groups are either similar or different. In multivariate procedures, however, researchers are often faced with dozens of variable combinations to interpret, and intuitive abilities may become taxed.

A third disadvantage of the multivariate approach is that it is easy to include so many variables that no sense can be made of the results. Although researchers are often tempted to include many potentially relevant variables in a multivariate study, guidelines must be established and followed to restrain this tendency. Finally, it should be noted that multivariate statistics are useful when a question calls for such analysis, but they are not a panacea for all research problems.

The bulk of this chapter is devoted to discussions of four of the most widely used multivariate statistical methods and examples of how they are used in mass media research. These discussions are simplified explanations which ignore many controversial aspects of each method for the sake of brevity. Readers interested in learning more about multivariate statistics should consult works listed in the "References and Suggested Readings" section at the end of the chapter.

The usefulness of multivariate statistics can be demonstrated by an example using the method called **multivariate analysis of variance** (MANOVA). (MANOVA is an extension of univariate ANOVA—see Chapter 11—and will be discussed in greater detail later in this chapter.) Assume researchers are interested in measuring the effects of television viewing and newspaper reading on academic exams in English, history, and economics. Their study design would look as follows:



This design offers four different situations to examine: "no paper/no TV"; "no paper/TV"; "paper/no TV"; and "paper/TV." More importantly, however, it allows the researchers to investigate the effects of the independent variables on all three exams simultaneously; univariate ANOVA would require three individual studies of each exam. It is clear that the multivariate procedure represents a significant savings in time, money, and resources.

All multivariate statistics are designed to reduce an original "test space," or group of data, into a minimum number of values or dimensions that describe the relevant information contained in the data; this is in accordance with the principle of parsimony. Thus, instead of using 20 dependent variables to describe a phenomenon, for example, a researcher might use multivariate methods to reduce the number to 3 summary variables (weighted linear combinations) that are nearly as accurate as the original 20. Data reduction involves little loss of information and makes the data easier to handle. It is an especially useful process in mass media research, where studies often deal with almost limitless numbers of variables (Cooley and Lohnes 1971).

Another quality of multivariate statistics is that they can serve both descriptive and inferential goals, often simultaneously. As Harris (1975) noted:

> On the descriptive side [multivariate statistics] provide rules for combining variables in an optimal way. What is meant by "optimal" varies from one technique to the next. On the inferential side, they provide a solution to the multiple comparisons problem.

Matrix Algebra

Multivariate procedures solve the problem of comparing multiple criterion variables by establishing *weighted linear combinations* of two or more such scores (see Chapter 11). These composite scores are represented as lines in space called **vectors.** Thus, multivariate statistics involve the manipulation of one or more vectors, usually in an attempt to predict an outcome or event through the principle of least squares (see Chapter 11). This manipulation of vectors requires a different type of analysis, involving what is known as *matrix algebra.*

Because all multivariate statistics deal with multiple measurements of data in multidimensional space, they are dependent upon matrix algebra. One does not require a complete knowledge of matrix algebra to understand multivariate statistics, but a familiarity with several frequently used terms is essential to follow the procedures described in the following sections.

A **scalar** is a single-digit number, such as 6, 9, or 7. A column of scalars is called a *column vector* and is denoted by a lowercase letter:

$$a = \begin{bmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ \cdot \\ a_n \end{bmatrix}$$

(The subscript represents the scalar's location in the column.) A row of scalars is referred to as a *row vector* and is denoted by a lowercase letter followed by a prime symbol:

$$a' = [a_1\ a_2 \ldots a_n]$$

A **matrix** is a two-dimensional array of scalars, having *p* rows and *n* columns, and is denoted by a capital letter:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdot & \cdot & \cdot & \cdot & a_{1n} \\ a_{21} & a_{22} & \cdot & \cdot & \cdot & \cdot & a_{2n} \\ a_{p1} & a_{p2} & \cdot & \cdot & \cdot & \cdot & a_{pn} \end{bmatrix}$$

The most common matrix used in multivariate statistics is the **intercorrelation** (or simply correlation) **matrix,** which is denoted as R. This matrix contains the coefficients of correlation between pairs of variables. For example, a 3 × 3 correlation matrix (the first number refers to the number of rows, the second to the number of columns) is used to display the relationship between three variables:

$$R = \begin{bmatrix} 1.0 & .64 & .29 \\ .64 & 1.0 & .42 \\ .29 & .42 & 1.0 \end{bmatrix}$$

This is a **square matrix:** the number of rows equals the number of columns. It contains the value 1.00 in the principal diagonal (top left to bottom right), since the correlation of a variable with itself is usually considered to be 100%. The term *usually* is used because in some multivariate models, it is assumed that even a correlation of a variable with itself includes some error and must hence be valued at less than 1.00.

A **diagonal matrix** is a square matrix whose elements equal zero except in the principal diagonal. If in addition the elements in the principal diagonal are all equal to 1, the array is known as an **identity matrix** and is denoted by a capital *I:*

$$I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

An identity matrix indicates that there are no correlations between the variables in the analysis except for the correlation of each variable with itself. Essentially, the identity matrix implies that the data are of poor quality. Identity matrixes are common in research studies where random numbers are used (usually in demonstrations of statistical procedures).

Associated with every square matrix is a single number that represents a unique function of the numbers contained in the matrix. This scalar function is known as a **determinant,** and is denoted as A (Cooley and Lohnes 1971). In addition, each square matrix has an associated characteristic equation that represents the information contained in the matrix. The results from the characteristic equation (when computed) reproduce the matrix from which the equation was developed. Several different values can be used to calculate a characteristic equation for a matrix; each of these includes a value or number known as an eigenvalue. Each eigenvalue has an associated eigenvector (a column of numbers); the eigenvalue is actually the sum of the squared elements in the eigenvector. In short, eigenvalues and eigenvectors are used to construct a formula that duplicates the information contained in a matrix. Each matrix has a number of different characteristic equations, but only one is appropriate for a particular type of analysis (see Cooley and Lohnes 1971; Rummel 1970 for more information).

Multivariate statistics involve two basic algebraic operations. The first is *partitioning* of a matrix, whereby the original matrix is divided into submatrixes for analysis. The second operation, *transposition*, involves changing the matrix columns to rows and the rows to columns for further analysis (Horst 1966). For example, the "transpose" of Matrix A is Matrix A':

$$A = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} \qquad A' = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$$

Four Multivariate Procedures

Although there are several multivariate statistics available, four methods appear to be used most often in mass media research. These are factor analysis, canonical correlation, discriminant analysis, and multivariate analysis of variance (MANOVA). Each method is discussed below in terms of how it works and what it can do in media research.

Factor Analysis

Factor analysis is a generic term for a variety of statistical procedures developed for the purpose of analyzing the intercorrelations *within* a set of variables. These relationships are represented by weighted linear combinations known as **factor scores** (or variates), which, in turn, are used in the development of constructs and theories. Factor analysis is divided into several different techniques; each technique is appropriate to a specific type of investigation. However, two major techniques are used most often in mass media research: *R-technique* and *Q-technique*. The R-technique is used to factor a set of variables collected at the same time from a number of individuals. The Q-technique is used to factor a number of individuals from variables collected at the same time from those individuals.

Each technique includes a variety of approaches. Some types of R-technique factor analysis include common factor analysis (CFA), principal components analysis (PC or PCA), and minimum residuals (Minres). The most often used procedure in the Q-technique is cluster analysis, which seeks to identify types of individuals, not groups of variables as in the R-technique.

Factor analysis is the most widely used multivariate statistic in mass media research, due to its flexibility. Some of the more common uses for factor analysis include (Rummel 1970):

1. Investigating patterns of variable relationships
2. Reducing data
3. Analyzing the structure of a phenomenon
4. Classifying or describing individuals, groups, or variables
5. Developing of measurement scales (see Chapter 2)
6. Testing hypotheses or research questions
7. Making preliminary investigations in new areas of research
8. Developing theories

As Rummel suggested, factor analysis is appropriate in any phase of research, from pilot studies to theory development. This is not true of other multivariate statistical procedures.

Researchers who use factor analysis assume that any group of variables has some inherent order that, once discovered, can make the description of a concept or construct less complicated. Thurstone (1947) noted that:

> A factor problem starts with the hope or conviction that a certain domain is not so chaotic as it looks. Factor analysis was developed primarily for the purposes of identifying principal dimensions or categories of mentality; but the methods are general, so that they have been found useful for other psychological problems and in other sciences as well. . . . [F]actor analysis is especially useful in those domains where basic and fruitful concepts are essentially lacking and where experiments have been difficult to achieve.

As Thurstone suggested, factor analysis is useful in all types of scientific research. This does not imply, however, that the results of a factor analysis in any particular area are necessarily *meaningful;* any matrix of variables can be factorially analyzed, but not all will yield scientifically useful or meaningful information (Gorsuch 1974). It is necessary to understand the purpose of factor analysis in order to determine its appropriateness for a specific mass media problem.

Factor analysis includes a wide variety of procedures. Because of the complexity of many of these methods, it is not possible to discuss each one in this chapter. This is left to multivariate statistics books and other more detailed texts.

The two most widely used forms of factor analysis are *principal components* and *principal factors* (also called common factor analysis). The methods are identical except for the initial step: the principal components model uses *unities* (1.00) in the principal diagonal of the correlation matrix, while the principal factors technique assumes that each variable's correlation with itself contains some degree of error and therefore cannot be equal to 100%. These correlations are called *communalities* and replace the unities in the principal diagonal of the original matrix. Communalities are a bit more complicated than unities, since they are only *estimates* of the correlations of variables with themselves. Although choosing the communality estimates to insert into the principal diagonal may sound simple, it is one of the most controversial decisions in factor analysis (Rummel 1970; Comrey 1973).

A typical factor analysis begins with the collection of data on a number of different variables (usually at the interval or ratio level). These quantified variables are then transformed via computer into a *product-moment correlation matrix*, the matrix that is usually used for factoring. To illustrate, consider the problem of determining which medium, or combination of media, a subject uses for news and information. A questionnaire is designed to collect this information with regard to five possible media sources: (1) radio; (2) television; (3) magazines; (4) newspapers; and (5) books. The correlation matrix developed from these data may look as shown in Figure 12.1. The matrix shows the correlations between all five variables. The principal diagonal divides the matrix into two parts, each a mirror image of the other. The diagonal itself is composed of unities, indicating that the method of principal components is being employed.

Note that although any matrix can be factored, even a matrix of random numbers, there are procedures that can help researchers determine whether a particular correlation matrix is valid. Two tests are

Figure 12.1 Example of correlation matrix (R)

	1	2	3	4	5
1	1.00	.92	.70	.38	.05
2	.92	1.00	.95	.71	.26
3	.70	.95	1.00	.88	.33
4	.38	.71	.88	1.00	.14
5	.05	.26	.33	.14	1.00

particularly useful: Bartlett's Sphericity Test and Kaiser's Measure of Sampling Adequacy (MSA). Both are used to determine the quality of the correlation matrix and to indicate whether the information is adequate for analysis.

In mathematical terms, factoring a matrix consists of extracting eigenvectors and their associated eigenvalues. These two sets of values are used to mathematically reproduce the correlation matrix. Eigenvectors are factor loadings, or numerical values from -1.00 to $+1.00$ indicating the amount of contribution each variable makes toward defining a factor. A factor loading is a quantified relationship—the farther its value is away from zero, the more relevant the variable is to the factor. The eigenvalues are used to determine which factors are relevant and should hence be analyzed (recall that an eigenvalue is computed by squaring and summing the elements in its eigenvector). One common procedure is to interpret only those factors with eigenvalues greater than 1.00 (although other methods are also used).

The eigenvectors and eigenvalues for the hypothetical study are displayed in Figure 12.2. This example shows that two factors "fell out" (were significant as determined by the eigenvalue cut-off of greater than 1.00) and may be used in explaining the media used for news and information. The eigenvalues are computed by squaring and summing each element in the eigenvector: $.85^2 + .66^2 + .37^2 + .52^2 + .29^2 = 1.65$.

The analysis does not stop with this initial extraction of factors; the initial factor loadings are generally too complex to be used for interpretation. Instead, researchers generally perform a second step, factor *rotation,* which essentially involves changing the multidimensional space in which the factors are located. Recall that eigenvectors (a vector of factor loadings) represent lines in space—a space visually constructed by x and y axes. The unrotated factor loadings are often in complex form; that is, there may be several complex variables where one variable loads significantly on more than one factor. Rotation attempts to clear this problem by changing the space in which the factor loadings are placed. The new factor loadings are mathematically equivalent to the original unrotated matrix, but more often

Figure 12.2 Unrotated factor matrix

	Factor 1	Factor 2	
Newspaper	.85	.53	Eigenvectors
Magazines	.66	−.09	
Books	.37	.73	
Radio	−.52	.34	
Television	.29	.63	
	1.65	1.33	Eigenvalues

represent a more meaningful set of factors—where the goal is to have each variable load significantly on only one factor. This additional ease of interpretation makes rotation appealing in behavioral research. A rotated factor matrix is shown in Figure 12.3.

The rotated factor matrix is mathematically equivalent to the unrotated matrix, as witnessed by identical eigenvalues, and the factors are now easier to interpret. The first step in interpretation is to identify those variables which are associated with one and only one factor. Here, it is evident that Variables 1, 2, and 3 "load" more heavily on Factor 1, while Variables 4 and 5 load more heavily on Factor 2. Three variables (1, 2, 3) are said to "define" Factor 1, and two variables (4, 5) define Factor 2.

The next step is to categorize the factors on the basis of the variables that define it. In this case, Factor 1 might be classified as "Print Media" and Factor 2 as "Electronic Media." In reality, however, Factor 2 might be eliminated altogether at this point, since it is defined by only two variables. It is customary when classifying factors to select only those factors that have at least three variables with significant loadings. Some researchers consider this practice controversial, but three significant loadings are necessary in order to establish the direction of the factor. If a factor has only two significant variable loadings—one positive and the other negative—one has no way of knowing if the factor itself is positive or negative; a third variable is required to provide the direction. Of course, this is not necessary when both variables are positive or negative; however, even so, a two-variable factor may be inadequate to explain the variance in the factor. As a general rule, then, it is best to consider only those factors that are defined by at least three variables.

Researchers must also consider which type of rotation to use. There are many procedures available, but the two used most often by behavioral researchers are *orthogonal* and *oblique* rotation. The names of these two procedures refer to the angles of the axes on which the data points (factor loadings) are located. Orthogonal rotation keeps the angles at 90° and assumes that the factors are not intercorrelated (orthogonal = uncorrelated). Oblique rotation allows the axes to take any angle in order to produce the most interpretable results; researchers

Figure 12.3 Rotated factor matrix

	Factor 1	Factor 2
Newspaper	.87	.15
Magazines	.66	−.03
Books	.61	.22
Radio	.17	.85
Television	.20	.74
	1.63	1.34

using this approach assume that the factors are correlated to some degree (oblique = correlated). The choice of a rotation method depends on the researcher's likes and dislikes as well as the purpose of the study.

Of the many possible uses of factor analysis mentioned earlier, the three that are actually most prominent in research studies are data reduction, the search for order in variable structures, and the exploration of uncharted phenomena. *Data reduction,* as noted earlier, is essential when investigating research problems that contain large numbers of variables. Factor analysis is often used as a preliminary narrowing device because it allows the selection of salient variables from a much larger group: it provides a simplification of a particular domain of variables by replacing them with a small number of hypothetical variates (weighted linear sums of the original variables, or factor scores). These variates can then be used in other analyses that employ different statistical methods without a substantial loss of information. For example, consider a 50-item questionnaire designed to measure attitudes toward commercials on television and their effects on buying habits. A factor analysis of the 50 variables would produce a substantially smaller number of representative variates, for factor scores, thus providing an opportunity for a much simpler explanation of the phenomenon in question (Comrey 1973).

Eastman's (1979) study provides a good demonstration of how factor analysis is used as a data reduction technique. The purpose of her study was to determine the factors affecting both the uses of television by and the lifestyles of a large group of subjects. To accomplish this, she distributed approximately 2,200 questionnaires to subjects in randomly selected neighborhoods in Fremont, Ohio. The usable questionnaires included responses from 1,795 subjects and included 250 items. On the basis of previous studies using smaller samples, Eastman was able to predict the number and type of factors that might be uncovered.

Her findings generally supported these hypotheses derived from previous analyses. Eastman concludes her article with a detailed discussion of these findings; the important point here, however, is that her work demonstrates data reduction. Consider the unruly task of analyzing 448,750 variables (1,795 × 250); such an undertaking would be impossible in any research study. By using factor analysis, however, Eastman was able to reduce the variables to 12 factor scores for each subject on the media use portion of the questionnaire and 14 factor scores on the lifestyle portion, thus allowing further analysis of the data. If such a reduction had not been performed, Eastman's hypotheses could never have been tested.

The *search for order* in a domain of variables via factor analysis is referred to by several different names: it can be said that factor analysis identifies variable patterns, dimensions of variable domains, underlying constructs, factor dimensions, or factor structures. Whatever terminology is used, the meaning is the same: factor analysis allows the identification from a large group of variables of a smaller number of composite variables that help order and define the phenomenon under study.

A construct such as "program success" may be defined by an infinite number of variables. It may be difficult, if not impossible, to intuitively determine which variables contribute significantly to the construct. Factor analysis, by reducing the number of variables, makes it easier to identify patterns and underlying structures. However, one factor analysis does not produce conclusive results concerning the composition of a construct; different samples, different factor analysis methods, and different variables must be used to verify that the initial results were not sample- or method-specific. That is, replication is necessary to ensure that results are not dependent upon some condition external to the relationship among the variables.

Finally, factor analysis can help to provide *explanations of previously unstudied phenomena.* Regardless of the research area in question, there are many concepts and constructs that have eluded investigation. One reason for this is that the concepts are obscured by large numbers of variables. An example is the question of what characterizes a successful television program. Television executives still do not know the answer to this; there are simply too many variables involved. Factor analysis can play a significant (albeit preliminary) role in solving problems of this nature by isolating salient variables from those that will add nothing to the accuracy of a prediction.

Factor analysis allows researchers to take a panoramic view of a variable domain and isolate important variables. However, it should be used only when a research project is well constructed and meets the prerequisite assumptions discussed above. It should *not* be used in an attempt to salvage a poorly planned study. As mentioned earlier, any matrix of variables can be factorially analyzed, but there is no guarantee that the results will be meaningful.

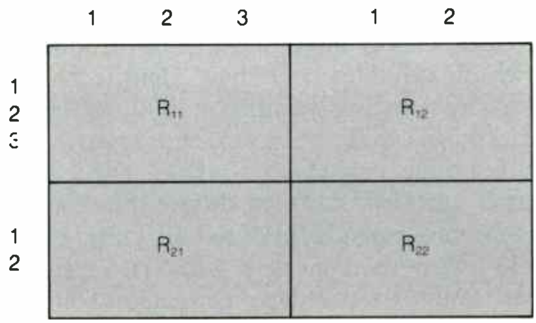
Canonical Correlation

Canonical correlation (R_c) is essentially multivariate extension of linear multiple regression: a group of independent variables is analyzed to predict multiple criterion variables. However, no distinction is actually made between independent and dependent variables; there are simply two *sets* of variables.

A canonical correlation begins with the formation of an intercorrelation "super-matrix" composed of both sets of variables. For example, if three variables are included in one set and two variables in the second set, the intercorrelation matrix is formed as Figure 12.4 shows.

Note that the correlation matrix is divided into four areas: R_{11} represents the intercorrelations among the elements of Set 1; R_{22} denotes intercorrelations among the elements of Set 2; and R_{12} and R_{21} are the cross-correlations between the elements of Sets 1 and 2.

Figure 12.4 R_c intercorrelation "super-matrix"



This "supermatrix" of intercorrelations provides the information necessary to compute the basic R_c relationship matrix by means of the formula:

$$R_c = [R_{22}{}^{-1}R_{21}R_{11}{}^{-1}R_{12}]$$

The canonical correlation matrix formed by the formula is a square matrix and therefore contains eigenvectors and eigenvalues (each unique solution is called a root; each root has a canonical correlation). A *pair* of eigenvectors is extracted from each variable set and then normalized; the resulting scalar values are the sets' beta weights. Finally, for each variable set, the subject's standardized raw scores are multiplied by the beta weight and the products summed. This process yields a pair of scores called *composite canonical variates* (weighted linear combinations of variables). The R_c is the product-moment correlation between these variates.

There are three basic values that a researcher interprets in canonical correlation analysis: (1) the canonical correlation for each root; (2) the canonical components; and (3) the redundancy index. The canonical correlation, since it is merely a product-moment correlation, is interpreted in the same way as any correlation value: the closer the value is to 1.00, the stronger the relationship is between the composite variates (canonical correlations cannot be negative). The R_c model allows the extraction of as many roots as there are variables in the smaller of the two sets (in the above example, the researcher could extract two unique roots), and each of these is orthogonal (uncorrelated) to all other roots in the analysis. This can provide several possibilities for interpretation (although not all of the roots may be statistically significant).

Interpretation of the individual variables in a canonical correlation involves analyzing their *canonical components*, that is, each variable's correlation with the canonical variate corresponding to its particular set. Thus, the canonical components are also correlations and interpreted as such: the components whose values are farthest from zero are considered most significant. However, all researchers do not agree as to what constitutes a significant component value; many consider .30 to be significant, while others use .35, .40, or some different value.

The **redundancy index** provides the direction in which the canonical results should be interpreted; that is, it determines whether the R_c is to be interpreted as "Set 1 given Set 2" or "Set 2 given Set 1," or, possibly, whether it should be interpreted in both directions. The redundancy index increases the interpretive value of canonical correlation by allowing researchers to interpret relationships within canonical components, as well as between the variable sets. In fact, relationships between sets should never be interpreted without first computing the redundancy index.

The formula for computing a redundancy index ($\bar{R}$) is simple: square each component value in a variable set, add the squares, divide the sum by the number of variables in the set, and then multiply this value by the canonical correlation squared:

$$\bar{R} = \frac{\Sigma(Rcc)^2}{M}\,(Rc^2)$$

where $\bar{R}$ = the redundancy index
R_{cc} = a canonical component
M = the number of variables in the set

The importance of these three values in canonical correlation analysis can be illustrated by an actual research problem. In a study of the 1968 presidential campaign, Wimmer (1976) investigated the research question: "What, if any, relationship exists between the mass media actually used for political information and the media considered to be most informative?" He took a survey of 3,100 subjects and obtained the results displayed in Figure 12.5.

First consider the four canonical correlations for the four roots. Each root has a value of at least .30, indicating that all may be interpreted unless the component loadings and redundancy index indicate otherwise (.30 is usually the cutoff value for interpretation of a canonical root; however, this depends on the nature of the study and the requirements established by the researcher). In most cases the roots in a canonical analysis do not all have a value of .30 or higher, but such values are common in studies using extremely large samples.

The second step is to examine the significant canonical components, which are considered to be those with absolute values equal to or greater than .30 (again, this limit is at the discretion of the researcher). Referring to Root 1, the variables in Set 1 (also known as the *left set*) show a high positive component value for television (.53) and a very low value for the variable "none" (−.30); the remaining variables have values too minimal to be considered significant. The left variable set of Root 1 thus suggests a degree of dichotomy: a great many subjects considered television the most informative medium for politi-

Figure 12.5 Canonical correlation example

Variables	Root 1 ($R_c = .85$)	Root 2 ($R_c = .45$)	Root 3 ($R_c = .31$)	Root 4 ($R_c = .31$)
		Left Set (Media Considered Most Informative)		
Newspaper	.28 (1.00)[a]	−.63 (−.29)	−.33 (−.07)	−.06 (−.03)
Radio	.02 (.35)	−.08 (.01)	.42 (.50)	−.87 (−.84)
Television	.53 (1.30)*	.53 (.62)	−.30 (.10)	.01 (.02)
Magazines	.16 (.54)	−.46 (−.31)	.64 (.72)	.46 (.45)
Paper & TV	.07 (.27)	−.02 (.03)	−.15 (−.10)	.05 (.05)
2 Combination	.06 (.22)	−.10 (−.05)	.13 (.17)	−.10 (.09)
3 Combination	.04 (.13)	−.09 (−.01)	.10 (.12)	−.02 (−.02)
People	.01 (.08)	.06 (.09)	−.05 (−.03)	.01 (.01)
None	−.30 (.50)*	.49 (.65)	.46 (.59)	.19 (.18)
		Right Set (Media Used for Political Information)		
Newspapers	.85 (.41)*	−.41 (−.82)	−.30 (−.99)	.02 (−.05)
Radio	.62 (.15)*	−.03 (.07)	.32 (.43)	−.70 (−1.10)
Magazines	.63 (.52)*	−.38 (.45)	.56 (.97)	.35 (.57)
Television	.90 (.52)*	.41 (1.10)	.02 (−.05)	.11 (.39)

[a]Canonical components are listed first with corresponding beta weights in parentheses.
*Indicates a significant canonical component value.

Root 1: $x^2 = 5{,}460.32$; df = 36; $p < .01$

Root 2: $x^2 = 1{,}358.26$; df = 24; $p < .01$

Root 3: $x^2 = 635.01$; df = 14; $p < .01$

Root 4: $x^2 = 313.93$; df = 6; $p < .01$

Redundancy Indexes

Left Set		Right Set	
Root 1:	.0470	Root 1:	.4308
Root 2:	.0273	Root 2:	.0258
Root 3:	.0117	Root 3:	.0126
Root 4:	.0111	Root 4:	.0154

cal information, but a smaller number indicated that none was most informative, possibly meaning that all were regarded as equally informative (the negative value means that those subjects who responded "television" as most informative did not respond "none"). All of the variables in Set 2 (the *right set*) surpassed the .30 cutoff value for Root 1, indicating that the subjects in the study used all of these media for political information.

At this stage, it is impossible to continue with the interpretation unless redundancy indexes are computed. That is, without redundancy figures, one can interpret only the relationships within sets, not those between sets. This point is repeated here because of its importance: many researchers incorrectly interpret the relationships between sets because they have neglected to compute redundancy indexes.

The general cutoff level for redundancy is .05. That is, a canonical root must account for at least 5% of the variance in the set before interpretation can continue. The redundancy indexes displayed in Figure 12.5 indicate that only Set 2 of the first root achieved a high enough value to qualify for further interpretation; this means that Root 1 is to be interpreted from the right set (media used), given the left set (media considered most informative).

The remaining three roots in the analysis did not receive significant redundancy indexes (even though their R_c values met minimum requirements); these roots must therefore either be eliminated from further discussion or interpreted for heuristic value only. This demonstrates how the redundancy index serves as a cross-validation for the canonical roots extracted from the analysis; it serves as a backup test for significance.

Applying this information to the data, the results show that those individuals who used all of the media for political information tended to feel that television was the most informative. In addition, a smaller group of individuals who used all the media felt that all were equally important. (For more information on canonical correlation, see the "References and Suggested Readings" at the end of the chapter.)

Discriminant Analysis

A frequent problem encountered in mass media research involves an interest in examining or predicting the attitudes or behavior of subjects who are members of a particular group. For example, a researcher might wish to examine the differences between subjects who subscribe to certain magazines or newspapers, or maybe wish to predict the characteristics of individuals occupying management level positions in the media. When these types of research situations are encountered, **discriminant analysis** can be a useful tool.

Discriminant analysis is a procedure whereby linear combinations of continuously scaled variables are derived from measurements made on groups of subjects. The model is used to define a vector that represents the variables for each group so that the separation between groups is maximized. In other words, a researcher using discriminant analysis uses all of the variables involved in the analysis to compute a weighted linear combination (variate) for each group, from which he or she can determine which of the variables are most helpful in separating or distinguishing the groups. These variables are known as the *discriminating variables* (Klecka 1980).

The discriminant analysis procedure is shown in graphic form in Figure 12.6. Here, a *bivariate* example is used (two groups). The two ellipses represent data *swarms* for each group, and the dot at the center of each ellipse denotes the group mean, or *centroid*. The two points where the ellipses intersect define a line, designated as "A." If a second line, "B," is constructed perpendicular to Line A, and the points from the intersection of Groups 1 and 2 are projected onto Line B, the overlap between the groups is smaller along Line B (indicated as a) than any other possible line (Cooley and Lohnes 1971). The discriminant analysis procedure attempts to define this *true* line along which the groups are maximally separated.

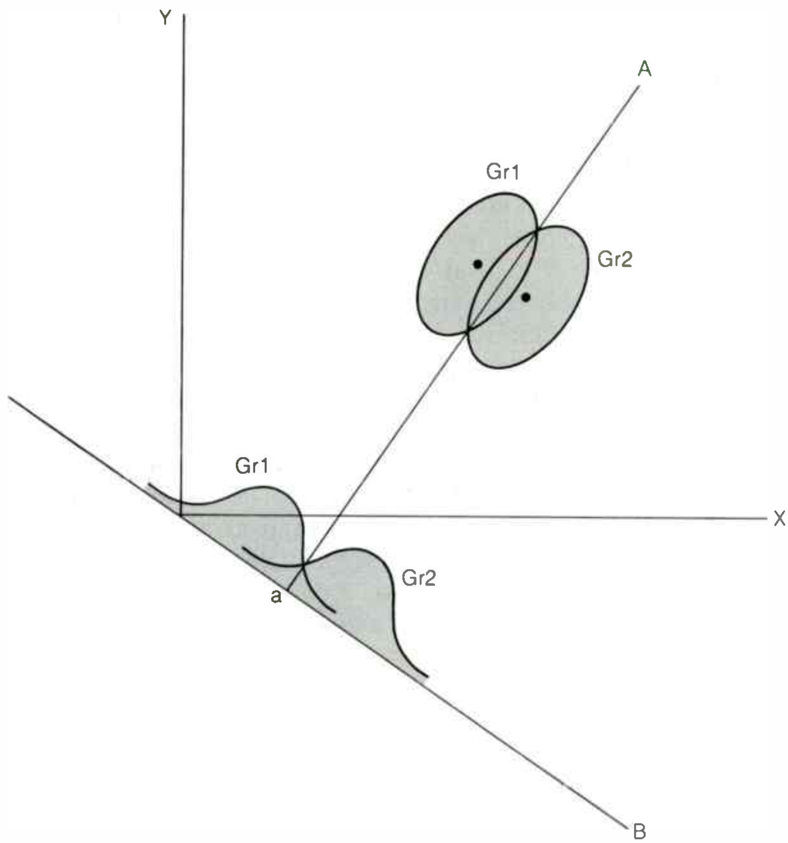
Discriminant analysis can serve two purposes, analysis and classification. The data are analyzed by means of statistical tests designed to measure the significance of the combined variables. Classification, which takes place after the data is analyzed, involves creating categorics of membership for the subjects involved in the study. A particular study may involve one or both of these procedures.

Discriminant analysis is closely related to factor analysis and canonical correlation in that each model extracts "factors" from a battery of variables. The factors, which are linear combinations in all models, are referred to as *discriminant functions*. Each discriminant function is orthogonal to all others in the analysis; that is, each function is an independent representation of the analysis in question. In discriminant analysis, the number of functions is one less than the number of groups involved, unless the number of variables in the analysis is smaller. In that case, the number of discriminant functions is equal to the number of original variables.

Interpreting a discriminant analysis involves examining the discriminant functions and the weights assigned to each of the variables used in forming those functions. These observations taken together allow researchers to analyze the nature of group differences. As in factor analysis, the dimensions represented by the discriminant functions may be susceptible to meaningful interpretation. However, even if the interpretation is not meaningful, parsimony is achieved by reducing the original space in which the group differences existed (Tatsuoka 1971).

Another procedure involved in discriminant analysis is plotting the centroids of the discriminant functions. Each group has a mean for the linear combination it creates; these means are plotted in order to

Figure 12.6 Discriminant analysis procedure



determine the nature of the distance between the groups—the plot provides a visual representation of group differences. Tests are also available to indicate whether the distance separating the groups is statistically significant (Tatsuoka 1970; Klecka 1980).

Haynes (1978) used discriminant analysis to test two hypotheses related to children's perceptions of comic and authentic violence in cartoons. A "comic cartoon" was defined as one that portrayed a violent act in a comical manner, where the victim suffered no true or lasting ill effects. An "authentic cartoon" was defined as one depicting a violent act as "true to life," with no comic effect intended. The two hypotheses were: (1) that children would perceive the violence in a comic cartoon as being more violent than that in an authentic cartoon; and (2) that females would perceive all of the cartoons as being more violent than would males.

Haynes' first step was to administer a 12-item questionnaire to a group of 120 children, asking them to describe how they felt about a cartoon they viewed in an experimental situation. These responses were factor analyzed and the factor scores used in the discriminant analysis to determine if group differences existed. (In this study, Haynes used discriminant analysis solely as a classification procedure, since he had no knowledge about group membership beforehand.)

Discriminant Function I (see Table 12.1) revealed that the males and females who viewed the comic cartoons (Group 1) were most clearly distinguishable from the males and females who viewed the authentic cartoons (Group 2) in terms of a function described as "perceived violence" (so named because of the nature of the most significantly weighted variables in the function). The second discriminant function showed that Group 1 was most different from Group 2 with regard to a function described as "acceptability of violence." The centroids for each group showed that no sex differences existed. However, Haynes did find that the comic cartoons were perceived as being more violent than the authentic cartoons, and that the comic violence was perceived as being more unacceptable than authentic violence.

Discriminant analysis is a useful research tool in all areas of mass media. The method is often used as a secondary phase of research, such as was done by Haynes: the researcher may conduct a factor analysis, or other statistical procedures that produce summary scores, and use these scores in a discriminant analysis to determine if group differences exist.

Table 12.1 Haynes' Discriminant Analysis of Male and Female Perception of "Comic" and "Authentic" Cartoon Violence

	Discriminant Function Coefficients	
	Function I	Function II
Perceived Violence	−.997	.079
Acceptability of Violence	.093	.994
	Centroids: DF I - Perceived Violence	
	Male-authentic	.35
	Female-authentic	.42
	Male-comic	−.48
	Female-comic	−.51
	Centroids: DF II - Acceptability	
	Male-authentic	.61
	Female-authentic	.65
	Male-comic	−.17
	Female-comic	−.12

Source Richard B Haynes, "Children's Perceptions of 'Comic' and 'Authentic' Cartoon Violence," *Journal of Broadcasting*, Winter 1978, p 68. Reprinted by permission

Multivariate Analysis of Variance

Multivariate analysis of variance (MANOVA) was introduced briefly at the beginning of this chapter to demonstrate the utility of multivariate statistics over univariate methods. As mentioned, MANOVA is an extension of the simple ANOVA model to situations involving more than one dependent variable. Specifically, MANOVA allows researchers to test the differences between two or more groups on multiple response data.

The distinctive feature of MANOVA is that the dependent variables are represented as a vector, or weighted linear combination, instead of as a single value, as in ANOVA. ANOVA involves testing for group differences along a continuum formed by the dependent variable, as is shown in Figure 12.7 (Cooley and Lohnes 1971).

The MANOVA model extends this idea by testing for group differences in a multidimensional space, as depicted in Figure 12.8.

The test used in MANOVA to determine the equality of centroids (as compared to the test for equality of group means in ANOVA) involves forming an F-ratio between the within groups' and total groups' *dispersion matrixes*. This concept is beyond the scope of an introductory text; suffice it to say that the procedure is similar to that used in ANOVA except that matrixes are used instead of sums of squares.

Lambert and his colleagues (1980) performed a study that illustrates how several dependent variables can be simultaneously analyzed using MANOVA. The project demonstrates that not only are time and resources saved, but also several characteristics of an individual can be considered simultaneously, thus providing a closer approximation of reality than could a univariate statistical approach.

Lambert and others were interested in examining the attitudes of consumers generally, and of the elderly in particular, toward the exploitation of opportunities to save money by substituting lower priced generic drugs for brand name products prescribed by their physicians. The authors used a multistage quota sampling procedure to select 510 respondents from four cities in Florida. The subjects

Figure 12.7 Simple ANOVA

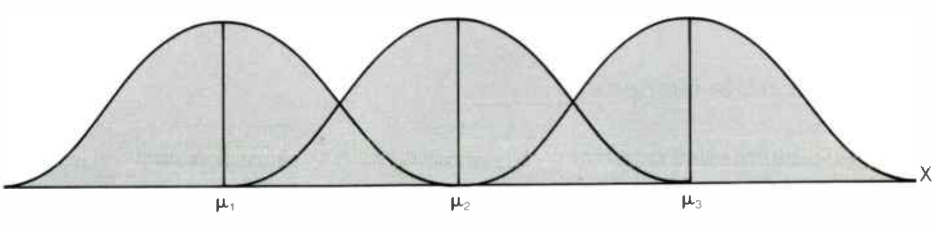
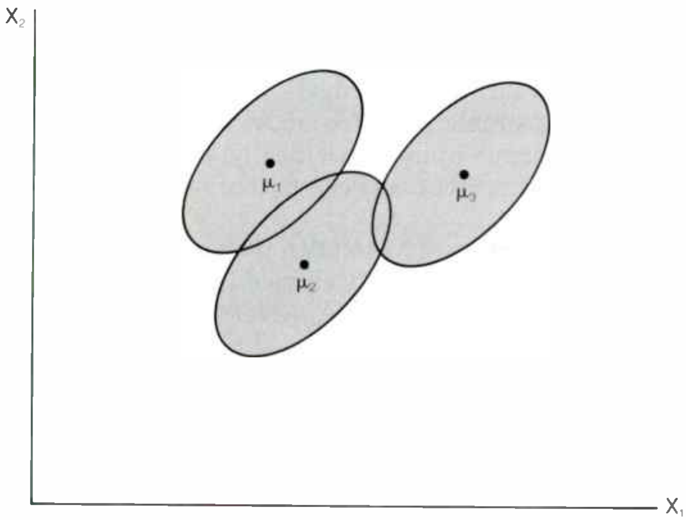


Figure 12.8 Simple MANOVA



completed a questionnaire containing 18 dependent variables. An interesting aspect of the study was that several of the respondents over age 65 refused to cooperate with the researchers for one reason or another (such as poor eyesight or fear of being victimized by salespeople). The study is an excellent example of the problems that can occur in some types of research studies.

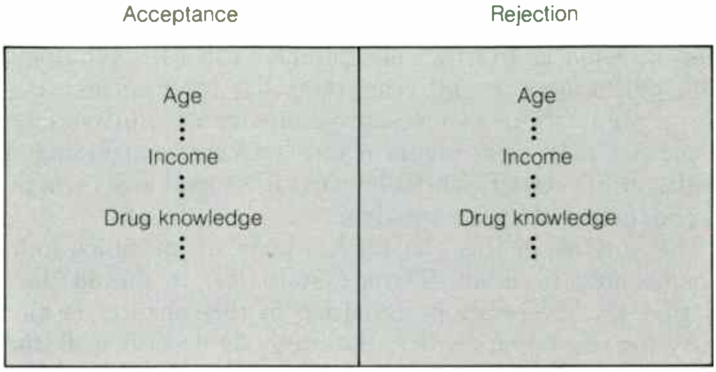
The authors divided the sample into two groups: those people who accepted the idea that generic drugs are equivalent to brand name drugs, and those who did not. These two groups were then compared with respect to their scores on the 18 variables, which included demographics, mobility, general drug knowledge, age, and income. The design is illustrated in Figure 12.9.

The MANOVA results indicated a significant difference ($p < .0001$) between the two groups on the 18 variables taken as a whole. The study allowed the authors to discuss the influence of many variables on consumers' attitudes toward generic drugs and to make recommendations concerning how drug education programs might help individuals, especially older people, take advantage of the lower cost generic drugs that are available.

Other Multivariate Methods

This chapter has described only four of the many multivariate statistical procedures available to mass media researchers. Several other statistics are appropriate for media researchers, some of which are described briefly in this section. Researchers interested in these areas

Figure 12.9 MANOVA example



are urged to consult the "References and Suggested Readings" at the end of the chapter.

Most researchers understand that human behavior and attitudes are comprised of a complex series of interacting variables, and that some variables are more important than others. In many situations, researchers may wish to define a series of steps involved in a decision or attitude and to "map out" the relationships between the relevant interacting variables. *Path analysis* is appropriate for this task.

Path analysis essentially creates a "causal path" to describe the relationship between the independent and the dependent variables (Duncan 1966). The method is particularly useful in situations involving decision-making processes or selections between several alternatives that might be followed in a given course of action. Path analysis does not *demonstrate* causality; rather, it is a method for tracing a set of causal assumptions made by the researcher.

Cluster analysis is closely related to factor analysis and involves grouping objects or persons together into "factors," or clusters, on the basis of their similarities and differences. A cluster analysis of variables is referred to as a V-analysis, and a cluster analysis of objects is called O-analysis (Tryon and Bailey 1970; Krippendorf 1980).

Multidimensional scaling (MDS) is used to analyze judgments of similarity between variables in such a way that the dimensionality of those judgments can be measured. The idea behind MDS is relatively simple. If each variable is thought of as a point in space, the problem becomes one of finding the smallest dimensionality of that space and eliminating errors in prediction (Torgerson 1958). In other words, one might conduct a research study of subjects' use of radio, television, and newspapers to predict product purchase behavior. The closer together that these three variables can be defined as points in space, the better will be the prediction of purchasing behavior.

Summary

The acceptance of multivariate statistics in mass media research has not been as rapid as in other disciplines, such as psychology. However, during the past several years there has been an increase in the number of publications of research studies using multivariate analysis. It appears that mass media researchers are recognizing the fact that multivariate statistics offer benefits in several areas where multiple dependent variables are involved.

The purpose of this chapter has been to introduce four of the most commonly used multivariate statistics. It should not be assumed that the information contained in this chapter is all that is necessary for researchers to use the methods described. Rather, it is strongly suggested that if a reader plans to perform any of these procedures, he or she should consult the appropriate publications listed in the "References and Suggested Readings" section at the end of the chapter.

Questions and Problems for Further Investigation

1. Design a simple mass media study using factor analysis. Explain why factor analysis is the most appropriate procedure to employ in this study.
2. Justify, in your own words, the use of multivariate statistical procedures in mass media research.
3. Why do you think the use of multivariate statistics has been slow to develop in mass media research?
4. How could canonical correlation be used to investigate demographic characteristics and their relationship to the popularity of certain television programs?
5. Review several issues of various mass media journals. How many articles use some type of multivariate analysis? What types of hypotheses or research questions do these articles investigate?

References and Suggested Readings

Cattell, R. B., ed. 1966. *Handbook of Multivariate Experimental Psychology.* Chicago: Rand McNally.

Comrey, A. L. 1973. *A First Course in Factor Analysis.* New York: Academic Press.

Cooley, W. W., and Lohnes, P. R. 1971. *Multivariate Data Analysis.* New York: Wiley.

Duncan, O. D. 1966. "Path Analysis: Sociological Examples." *American Journal of Sociology* 72:1–16.

Eastman, S. T. 1979. "Uses of Television Viewing and Consumer Lifestyles: A Multivariate Analysis." *Journal of Broadcasting* 23:491–500.

Gorsuch, R. L. 1974. *Factor Analysis.* Philadelphia: W. B. Saunders.

Harris, R. 1975. *A Primer of Multivariate Statistics.* New York: Academic Press.

Haynes, R. B. 1978. "Children's Perceptions of 'Comic' and 'Authentic' Cartoon Violence." *Journal of Broadcasting* 22:63–70.

Horst, P. 1966. "An Overview of the Essentials of Multivariate Analysis Methods." *Handbook of Multivariate Experimental Psychology,* ed. by R. B. Cattell. Chicago: Rand McNally.

Krippendorf, K. 1980. "Clustering." *Multivariate Techniques in Human Communication Research,* ed. by P. R. Monge and J. N. Cappella. New York: Academic Press.

Klecka, W. R. 1980. *Discriminant Analysis.* Beverly Hills: Sage.

Lambert, Z. V., Doering, P., Goldstein, E., and McCormick, W. 1980. "Predisposition Toward Generic Drug Acceptance." *The Journal of Consumer Research* 7:14–23.

Rummel, R. J. 1970. *Applied Factor Analysis.* Evanston: Northwestern University Press.

Tatsuoka, M. M. 1970. *Discriminant Analysis.* Champaign, Ill.: Institute for Personality and Ability Testing.

Tatsuoka, M. M. 1971. *Multivariate Analysis.* New York: Wiley.

Thurstone, L. L. 1947. *Multiple Factor Analysis.* Chicago: University of Chicago Press.

Torgerson, W. 1958. *Theory and Methods of Scaling.* New York: Wiley.

Tryon, R. C., and Bailey, D. E. 1970. *Cluster Analysis.* New York: McGraw-Hill.

Tucker, R. K. 1982. *Basic Multivariate Research Models.* San Diego: College Hill Press.

Wimmer, R. D. 1976. "A Multivariate Analysis of the Uses and Effects of the Mass Media in the 1968 Presidential Campaign." Unpublished doctoral dissertation, Bowling Green State University.

PART 4

Research Applications

CHAPTER 13

Research in the Print Media

Methodologies used to study the print media are similar to those employed in most areas of research; academic and commercial research organizations often employ content analysis, experiments, focus groups, and surveys, among other procedures, to study newspapers and magazines. Print media research, however, tends to be more narrowly focused and more oriented toward practical application than in other fields. This chapter provides a brief overview of the most common types of studies in newspaper and magazine research, with a special emphasis on the research most likely to be conducted by advertiser-supported publications.

This chapter does not attempt to deal with *basic market studies* and *advertising exposure studies.* A basic market study provides a statistical portrait of the potential readers of a newspaper or magazine in terms of their demographic or psychographic characteristics. This market research technique is described more fully by Ferber (1974). Advertising exposure studies (also called reader traffic studies) are conducted to determine which advertisements are noticed or read by a publication's audience. More information on these studies is contained in Chapter 15.

Background

Research dealing with magazines and newspapers was one of the first areas of mass communication research to be developed. The initial interest in such research came from colleges and universities. In 1924, *The Journalism Bulletin* was first published by the Association of American Schools and Departments of Journalism. The first issue of that journal contained an article by William Bleyer, entitled "Research Problems and Newspaper Analysis," in which he presented a list of possible research topics in journalism. Among them were: (1) the effects of form and typography on the ease and rapidity of newspaper reading; (2) the effects of newspaper content on circulation; and (3) analysis of newspaper content. Bleyer's article was remarkably accurate in predicting the types of studies that would characterize newspaper and magazine research in the coming years.

Much of the content of early print media research was qualitative. *Journalism Quarterly* was founded in 1928 as the successor to *The Journalism Bulletin;* its first volume contained articles dealing with press law, history, international comparisons, and ethics. By 1930, however, quantitative research began to make its appearance in this academic journal: an article published in March of that year surveyed the research interests of those currently working in the newspaper and magazine field and found the most prevalent type of study to be the survey of reader interest in newspaper content. The June 1930 issue contained an article by Ralph Nafziger, "A Reader Interest Survey of Madison, Wisconsin," which served as the prototype for literally hundreds of future research studies.

The 1930s also saw the publication of many studies designed to assess the results of print media advertising. This led to studies in applied research, and several publications began sponsoring their own readership surveys. By and large, however, the results of these studies were considered proprietary.

As the techniques of quantitative research became more widely known and adopted, newspaper and magazine research became more empirical. The growth of this trend was first recognized by Wilbur Schramm (1957) in an article in *Public Opinion Quarterly* that reviewed 20 years of research as reported in *Journalism Quarterly.* Schramm found that only 10% of the 101 articles published between 1937 and 1941 concerned quantitative analyses. By 1952–56, nearly half of the 143 articles published were quantitative, a fivefold increase in only 15 years. The reasons for this growth, according to Schramm, were the growing availability of basic data, the development of more sophisticated research tools, and the increase in institutional support for research.

By 1960, newspapers and magazines were competing with television as well as radio for audience attention and advertiser investment. This situation greatly spurred the growth of private sector research. The Bureau of Advertising of the American Newspaper Publishers Association (now called the Newspaper Advertising Bureau) began conducting studies on all aspects of the press and its audience. In the 1970s, it founded the News Research Center, which publishes a report designed to present the results of research to editors. The Magazine Publishers Association also began to sponsor survey research at about this same time. The continuing interests of academics in print media research led to the creation of the *Newspaper Research Journal* in 1979, a publication devoted entirely to research that has practical implications for newspaper management.

By 1980, print media research was being conducted by universities, colleges, in-house research organizations, professional associations, and commercial research firms. The majority of this research today is quantitative, although qualitative articles still appear frequently in print media research journals.

Types of Print Media Research

Newspaper and magazine researchers conduct four basic types of studies: (1) readership studies, (2) circulation studies, (3) studies of typography and makeup, and (4) readability studies. Most of their research falls into the first category; circulation studies rank second but are far less numerous, while relatively few studies fall into the last two categories. Consequently, this chapter will concentrate on the various forms of readership research.

Readership Research

Many readership studies were done in the United States in the years immediately preceding and following World War II. The George Gallup organization was a pioneer in the development of the methodology of these studies: a personal interview in which respondents were shown a copy of a newspaper and asked to identify the articles they had read. The most complete survey of newspaper readership was the American Newspaper Publishers Association's *Continuing Studies of Newspapers,* which involved over 50,000 interviews with readers of 130 daily newspapers between 1939 and 1950 (Swanson 1955).

Readership research became important to management during the 1960s and 1970s, as circulation rates in metropolitan areas began to level off or decline and editors and publishers became concerned with holding the interests of their readers. They began more than ever to depend on surveys for the detailed audience information they needed to shape the content of a publication.

Today, research into newspaper readership is composed primarily of five types of studies. The first, called a **reader profile,** provides a demographic summary of the readers of a particular publication. For example, a profile of the audience of a travel-oriented magazine might disclose that the majority of the readers earn more than $40,000 a year, are 25–34 years old, hold college degrees, possess six credit cards, and travel at least three times a year. This information can be used to focus the publication, prepare advertising promotions, and increase subscriptions.

A second type of newspaper readership study is the **item-selection study,** which is used to determine who reads specific parts of the paper. The readership of a particular item is usually measured by means of **aided recall,** whereby the interviewer shows a copy of the paper to the respondent to find out which stories the respondent remembers reading or looking through. A variation on this technique is to have the interviewer preselect those items for which readership data are to be gathered and to ask subjects if they remember only those items.

Due to the expense involved in conducting personal interviews, some researchers now use phone interviews to collect readership data. This method requires that calls be made on the same day that the issue of the paper is published. When the interviewer reaches a reader of that day's paper, he or she asks the respondent to bring a copy of the paper to the phone. The interviewer, who also has a copy of the paper, then goes over each page with the respondent, who identifies those particular items he or she has read. Although this method saves money, it excludes from study those readers who do not happen to have a copy of the paper handy.

Another money-saving technique is to mail respondents a self-administered readership survey. Hvistendahl (1977) described two variations of this type of study. In the "whole copy" method, a sample of respondents receives an entire copy of the previous day's paper in the mail, along with a set of instructions and a questionnaire. The instructions direct the respondents to go through the newspaper and mark each item they have read by drawing a line through it. A return envelope with postage prepaid is provided. In the "clipping" method, the procedure is identical except that respondents are mailed clippings of certain items rather than the whole paper. To save postage fees, the clippings are pasted up on pages, reduced 25%, and reproduced by offset. Hvistendahl reported a 67% return rate using this method with only one follow-up postcard. He noted that the whole copy and clipping methods produced roughly equivalent results, although readership scores on some items tended to be slightly higher when clippings were used. A comparison of the results of these self-administered surveys with the results of personal interviews also indicated a basic equivalence.

Recently Stamm, Jackson, and Jacoubovitch (1980) suggested a more detailed method of item-selection analysis which they called a **tracking study.** They supplied their respondents with a selection of colored pencils and asked them to identify which parts of an article (headline, text, photo, outline) they had read, using a different colored pencil each time they began a new *reading episode* (defined as a stream of uninterrupted reading). The results showed a wide degree of variability in the readership of the various elements that made up an item: For one story, 27% of the subjects had read the headline, 32% the text, and 36% the outline. There was also variation in the length and type of articles read per reading episode.

The unit of analysis in an item-selection study is a specific news article (such as the story on page one dealing with a fire) or a specific content category (such as crime news, sports, obituaries). The readership of items or categories is then related to certain audience demographic or psychographic characteristics. For example, Larkin and Hecht (1979) found that readers of nonmetropolitan daily papers read news about local events the most and news about national events the least. Schwartz, Moore, and Krekel (1979) constructed a psychographic profile of frequent newspaper readers that placed each reader into one of four categories—young optimists, traditional conservatives, progressive conservatives, and grim independents—on the basis of which sections of the paper they tended to read. Young optimists, for example, were heavy readers of astrology columns, housing ads, and the classified section; in contrast, grim independents were heavy consumers of sports and business news. In another readership study, Lynn and Bennett (1980) divided their sample according to residence in urban, rural, or farming areas. Their survey found that there was little difference in the type of news content read by farm and rural dwellers, but that urban residents were more likely to read letters to the editor, society items, and local news.

The third type of newspaper readership research is called the **reader–non-reader study.** This type of study can be conducted via personal, telephone, or mail interviews with minor modifications. A problem in conducting reader–non-reader studies is establishing an operational definition for the term "non-reader." In some studies, a non-reader is determined by a "no" answer to the question, "Do you generally read a newspaper?" Others have used the more specific, "Have you read a newspaper yesterday or today?" (the rationale being that respondents are more likely to admit they haven't read a paper today or yesterday than that they never read one). A third form of this question uses multiple response categories. Respondents are asked, "How often do you read a daily paper?" and are given five choices of response: "very often," "often," "sometimes," "seldom," and "never." Non-readers are defined as those people who check the "never" response or as those who indicate either "seldom" or "never." Obviously, the form of the question has an impact on how many people are classified as non-readers. The largest percentage of non-readers generally occurs when researchers ask, "Have you read a newspaper today or yesterday?" (Penrose et al. 1974); the smallest number is obtained by requiring a "never" response to the multiple response question (Sobal and Jackson-Beeck 1981).

Once the non-readers are identified, researchers typically attempt to describe them by means of traditional demographic variables. For example, Penrose and others (1974) found non-reading to be related to low education, low income, and nonurban residence. Sobal and Jackson-Beeck (1981) reported that non-readers tend to be older, to have less education and lower incomes, and to have more often been widowed or divorced than readers. And Bogart (1981) concluded that non-readers are less likely to have voted in the last presidential election and to believe that their opinions had an impact on local government.

Several non-reader studies have attempted to identify the reasons for not reading the newspaper. The data for these subjects have generally been collected by asking non-readers to tell the interviewer in their own words why they don't read. Responses are then analyzed and the most frequent reasons reported. Poindexter (1978) found that the three reasons named most often by non-readers were lack of time, preference for another news medium (especially TV), and cost. Bogart (1981) identified four reasons: depressing news, cost, lack of interest, and inability to spend sufficient time at home.

The **uses and gratifications study** is a fourth type of newspaper readership research; it is used to determine the motives that lead to newspaper reading and the personal and psychological rewards that result from it. The methodology of the uses and gratifications study is straightforward: respondents are presented with a list of possible uses and gratifications and are asked whether any of these are the motives behind their reading. For example, a reader might be presented with the following question:

> Here is a list of some things people have said about why they read the newspaper. How much do you agree or disagree with each statement?
>
> 1. I read the newspaper because it is entertaining.
>   _____ agree   _____ not sure   _____ disagree
> 2. I read the newspaper because I want to kill time.
> 3. I read the newspaper to keep up to date with what's going on around me.
> 4. I read the newspaper to relax and to relieve tension.
> 5. I read the newspaper so I can find out what other people are saying about things that are important to me.

The responses are summed and an average score for each motivation item is calculated.

Several studies have taken this approach to explain readership. For example, McCombs (1979) found three primary psychological motivations for reading newspapers: the need to keep up to date, the need for information, and the need for fun. Reading for information seemed to be the strongest factor. Similarly, Weaver, Wilhoit, and Reide (1979) found that the three motivations most common in explaining general media use are the need to keep tabs on what is going on around one, the need to be entertained, and the need to kill time. The authors also noted differences among demographic groups as to which of these needs were best met by the newspaper. For example, young males, young females, and middle-aged males were most likely to say they used a newspaper to satisfy their need to keep tabs on things, but they preferred other forms of media for entertainment and killing time.

The final area of newspaper readership research involves **editor–reader comparisons.** In this type of study, a group of editors is questioned about a certain topic, and their answers are compared to those of their readers to see if there is any correspondence between the two groups. Bogart (1981) presented two examples of such research. In one study, a group of several hundred editors was asked to rate 23 attributes of a high-quality newspaper. The editors ranked "high ratio of staff-written copy to wire service copy" first; "high amount of nonadvertising content" second; and "high ratio of news interpretations . . . to spot news reports" third. The same list of attributes was then given to a sample of readers. The three top attributes according to the editors were ranked seventh, eleventh, and twelfth, respectively, by the readers, who rated "presence of an action line column" first; "high ratio of sports and feature news to total news" second; and "presence of a news summary" and "high number of letters to the editor per issue" as a tie for third. In short, there was little congruence between the two groups in the way they perceived the attributes of a high-quality newspaper.

In a related study, Bogart gave readers an opportunity to design their own newspaper. Interviewers presented a sample of readers with 34 subjects and asked them how much space they would give to each if they were designing a paper tailor-made to their own interests. Major categories of news were omitted from the listings because they were topics over which editors have little control. Once the results were tabulated, the contents of a sample of newspapers were analyzed to see if the space allocations made by editors matched the public's preferences. The resulting data indicated that readers want more of certain content than they are getting (consumer news; health, nutritional and medical advice; home maintenance; travel) and that they were getting more of some topics than they desired (sports news; human interest stories; school news; crossword puzzles; astrology).

Magazine readership surveys are fundamentally similar to those conducted for newspapers, but the two tend to differ in their particulars. Some magazine research is done by personal interview; the respondent is shown an actual copy of the magazine under study and is asked to rate each article on a four-point scale ("read all," "read most," "read some," or "didn't read"). The mail survey technique, also frequently used, involves sending a second copy of the magazine to a subscriber shortly after his or her regular copy has been mailed; instructions on how to mark the survey copy to show readership are included. For example, the respondents might be instructed to mark with a check those articles that they scanned, to draw an "X" through articles read in their entirety, and to underline articles that were only partly read.

Many magazines maintain *reader panels* of 25–30 people who are selected to participate for a predetermined period of time. All feature articles appearing in each issue of the magazine are sent to these panel members, who rate each article on a number of scales, including interest, ease of readership, and usefulness. Over time, a set of guidelines for evaluating the success of an article is drawn up, and future articles can be measured against that standard. The primary advantage of this form of panel survey is that it can provide information about audience reactions at a modest cost.

Another procedure that is peculiar to magazine research is the **item pretest** (Haskins 1960). A random sample of magazine readers is shown an article title, a byline, and a brief description of the content of the story. Respondents are asked to rate the idea on a scale from 0 to 100, where "100" represents "would certainly read this article" and "0" represents "would not read this article." The average ratings of the proposed articles are tabulated as a guide for future editorial decisions. Note that this technique can be used in personal interviews or with a mail survey with little variation in approach. Haskins also reported that there was a positive correlation between scores obtained using this technique and those determined by postpublication readership surveys.

Circulation Studies

The term *circulation research* is applied to two different forms of newspaper and magazine studies. One form attempts to measure circulation in terms of the overall characteristics of a particular market. For example, researchers might use this approach to determine the proportion of households in a given market that are reached by a particular newspaper or the circulation pattern of a magazine among certain demographic groups or in specific geographic areas. Tillinghast (1981), for example, analyzed changes in newspaper circulation in four regions of the country and found that the greatest decrease had occurred in the East and the South. He also reported that the degree of urbanization in a region was positively related to circulation. Nunn (1979) studied the relationship between circulation and separate editorial management for the morning and evening editions of certain newspapers. He found that papers with separate morning and evening editors had a higher circulation ratio than papers that used the same editor for both editions; this difference was most pronounced in large metropolitan markets. In a study of 69 Canadian daily newspaper markets, Alperstein (1980) discovered that newspaper circulation was positively related to the proportion of reading households within the newspaper's home city. In addition, daily newspaper circulation was found to be inversely related to weekly newspaper circulation. Stone (1978) provides a review of other market characteristics that have been employed in analyzing newspaper circulation.

Some researchers have used computer models with this form of circulation research. *Playboy*, for example, collected data for 52 issues of its publication that included number of copies sold, cover price, current unemployment statistics, dollars spent on promotion, number of days on sale, editors' estimates of the cover, number of full-page displays, and several other variables. These figures were then subjected to a regression analysis to determine how each factor was related to total sales. Interestingly, the number of copies distributed, the number of days an issue was on sale, and the cover rating proved to be good predictors, while the amount of money spent on promotion was found to have little impact on sales. This kind of analysis is generally limited to large publication companies because of the expense involved in data collection and using the computer for data storage and analysis.

The other type of circulation research uses the individual reader as the unit of analysis to measure the effects of certain aspects of delivery and pricing systems on reader behavior. For example, McCombs, Mullins, and Weaver (1974) studied why people cancel their subscriptions to newspapers. He found that the primary reasons had less to do with content than with circulation problems, such as irregular delivery and delivery in an unreadable condition. Magazine publishers often conduct this type of circulation research by drawing samples of subscribers in different states and checking on the delivery dates of their publication and its physical condition when received. Other publications contact subscribers who don't renew to determine what can be done to prevent cancellations. In recent years, several newspapers have researched the effects of price increases on their circulation. Studies have even been conducted to ascertain why some people don't pay their subscription bills when they are due. In short, this form of circulation research investigates the effect on readership or subscription rates of variables that are unrelated to a publication's content.

Typography and Makeup

Another type of study that is unique to print media research measures the effects of news design elements, specifically typeface and page makeup, on readership and reader preferences. By means of this approach, researchers have tested the effects of dozens of different typography and makeup elements, including: (1) amount of white space; (2) presence of paragraph headlines; (3) size and style of type; (4) variations in column width; and (5) use of vertical vs. horizontal page makeup.

The experimental method (Chapter 5) is used most often in typography and makeup studies. Subjects are typically assigned to one or more treatment groups, exposed to an experimental stimulus (typically in the form of a mock newspaper or magazine page), and asked to rate what they have seen according to a series of dependent variable measures.

Among dependent variables that have been rated by subjects are the informative value of a publication, interest in reading a publication, the "image" of a page, recall of textual material, readability, and general preference for a particular page. A common practice is to measure these variables by means of a semantic differential rating scale. For example, Siskind (1979) used a nine-point, 20-item differential scale with such adjective pairs as "informative–uninformative," "unpleasant–pleasant," "easy–difficult," "clear–unclear," "messy–neat," "bold–timid," and "passive–active." She obtained a general reader preference score by having subjects rate a newspaper page and summing their responses to all 20 items. Other studies have measured reader interest by using the rating scale technique or the 0–100 "feeling thermometer" (Chapter 7). Comprehension and recall are typically measured by a series of true-false or multiple-choice questions on the content that is being evaluated.

Haskins and Flynne (1974) conducted a typical design study to test the effects of different typefaces on perceived attractiveness of and reader interest in the women's section of a newspaper. They hypothesized that some typefaces would be perceived as more "feminine" than others and that headlines in such typefaces would create more reader interest in the page. The authors showed an experimental copy of a newspaper prepared specially for the study to a sample of 150 female heads of households: one subsample saw a paper where the headlines in the women's section had been printed in Garamond Italic (a typeface that experts had rated as being "feminine"), while a second group saw the same page with Spartan Black headlines (considered to be a more "masculine" typeface). A third group served as a control and saw only the headlines typed on individual white cards. The subjects were asked to evaluate each article for reading interest. Additionally, each woman was shown a sample of ten different typefaces and asked to rate them on a semantic differential scale with 16 adjective pairs.

The researchers discovered that typeface had no impact on reader interest scores. In fact, the scores were about the same for the printed headlines as they were for those typed on white cards. Analysis of the typeface ratings revealed that readers were able to differentiate between typefaces; Garamond Italic was rated as the second most "feminine" typeface while Spartan Black was rated most "masculine," thus confirming the judgment of the expert raters.

Studies of page layout have been used to help magazine editors make decisions about the mechanics of editing and makeup. Click and Baird (1979) have provided a summary of the more pertinent research in this area. A few of their conclusions are listed here to illustrate the types of independent variables that have been studied:

1. Large illustrations attract more readers than small ones.
2. Unusually shaped pictures irritate readers.
3. A small amount of text and a large picture on the opening pages of an article increases readership.
4. Readers do not like to read type set in italics.
5. For titles, readers prefer simple, familiar typefaces.
6. Readers and graphic designers seldom agree as to what constitutes superior type design.
7. Roman type can be read more quickly than other typefaces.

Readability

Simply defined, **readability** is the sum total of all the elements and their interactions that affect the success of a piece of printed material. "Success" is measured by the extent to which readers understand the piece, are able to read it at an optimum speed, and find it interesting (Dale and Chall 1948).

Several formulas have been developed to objectively determine the readability of text. One of the best known is the **Flesch** (1948) **reading ease formula,** which requires the researcher to: (1) systematically select 100 words from the text; (2) determine the total number of syllables in those words (*wl*); (3) determine the average number of words per sentence (*sl*); and (4) compute the following equation:

$$\text{Reading Ease} = 206.835 - .846wl - 1.015\,sl$$

The score is then compared to a chart that provides a description of style (such as "very easy") or a school grade level for the potential audience.

Another measure of readability is the **Fog Index,** which was developed by Gunning (1952). To compute the Fog Index, researchers must: (1) systematically select samples of 100 words each; (2) determine the mean sentence length by dividing the number of words by the number of sentences; (3) count the number of words with three or more syllables; (4) add the mean sentence length to the number of words with three or more syllables; and (5) multiply this sum by .4. Like the Flesch index, the Gunning formula suggests the educational level required for understanding a text. The chief advantages of the Fog Index are that the syllable count and the overall calculations are simpler to perform.

McLaughlin (1969) proposed a third readability index called **SMOG Grading** (Simple Measure of Gobbledygook). The SMOG Grading is quick and easy to calculate: the researcher merely selects ten consecutive sentences near the beginning of the text, ten from the middle, and ten from the end; counts every word of three or more syllables; and takes the square root of the total. The number thus obtained represents the reading grade that a person must have reached to understand the text. McLaughlin's index can be quickly calculated using a small, easily measured sample. Although the procedure is related to that for the Fog Index, it appears that the SMOG grade is generally lower.

Taylor (1953) developed yet another method for measuring readability called the **Cloze procedure.** This technique departs from the formulas listed above in that it does not require an actual count of words or syllables. Instead, the researcher: (1) chooses a passage of about 250–300 words from the text; (2) deletes every fifth word from a random starting point and replaces it with a blank; (3) gives the passage to subjects and asks them to fill in the blanks with what they think are the correct words; and (4) counts the number of times the blanks are replaced with the correct words. The number of correct words or the percentage of correct replacement constitutes the readability score for that passage. The paragraph below is a sample of what a passage might look like after it has been prepared for the Cloze procedure:

> The main stronghold of the far left ____ to be the large ____ centers of North Italy. ____ is significant, however, that ____ largest relative increase in ____ leftist vote occurred in ____ areas where most of ____ landless peasants live—in ____ and south Italy and ____ Sicily and Sardinia. The ____ had concentrated much of ____ efforts on winning the ____ of those peasants.

Nestvold (1972) found that Cloze procedure scores were highly correlated with readers' own evaluations of content difficulty. The Cloze procedure was also found to be a better predictor of such evaluations than several other common readability tests.

Although they are not used extensively in print media research, readability studies can provide valuable information. For example, Fowler and Smith (1979), using samples from 1904, 1933, and 1965, found that text from magazines had remained constant in readability while text from newspapers had fluctuated. For all years studied, magazines were easier to read than newspapers. Hoskins (1973) analyzed the readability levels of Associated Press and United Press International wire copy and found that both services scored in the "difficult" range; the Flesch indexes indicated that a 13th- to 16th-grade education was necessary for comprehension.

Print Media Research: An Overview

In 1977, the American Newspapers Publishers Association (ANPA) began to compile, review, and index research studies and private reports dealing with newspaper readership and circulation. The first summary of the project was reported by McCombs (1977) and updated in 1979 (Poindexter 1979). The research literature was categorized according to 16 major dependent variables that have been studied by newspaper researchers. The most frequent research method was item-selection, followed by reader–non-reader studies and studies measuring the amount read in the newspaper. Figure 13.1 lists the dependent variables identified in the ANPA summary.

Readers who are interested in pursuing newspaper research studies can make use of the ANPA computerized bibliography service, which indexes the past 30 years of readership studies. The index contains six pieces of information about each study: (1) independent variables; (2) dependent variables; (3) date of publication; (4) source of the study; (5) whether the report was public or confidential; and (6) whether the study was a major or a minor one. Researchers can request either a bibliography of relevant studies or a set of abstracts. The computerized bibliography is housed at the ANPA News Research Center at Syracuse University. There is a nominal charge to cover the cost of computer time used in making the search.

The amount and type of research done in the magazine industry is tabulated by *Folio* magazine. *Folio* periodically surveys about 600 top magazine executives about the research, if any, that is conducted by or for their publications. Specifically, *Folio*'s surveys are designed to determine how many and what kinds of magazines use research, what types of research they use, how valuable it is to them, and how

Figure 13.1 Newspaper research interest: focal questions

Type of Study	Total Number of Research Findings 1950–1979
Item-Selection	179
Reader–Nonreader	75
Amount Read	34
Selection of Particular Paper	33
Frequency of Reading	31
Selecting Newspaper over Other Media	31
Selecting Newspaper or TV for News	30
Subscribing vs. Nonsubscribing	23
Circulation Penetration	22
Other Studies	96

Source: Paula Poindexter, "Newspaper Readership and Circulation, AN UPDATE: 1977–79," American Newspaper Publishers Association, ANPA NEWS RESEARCH REPORT, No. 22, September 28, 1979. Reprinted by permission.

much they spend on it. The most recent of these surveys was conducted by Love (1981); the results included the following observations:

1. Business magazines conduct research more often than consumer magazines: 70% of all business magazines responding to the survey conducted research, compared to 60% of consumer magazines.
2. The income of a magazine is related to its research effort: virtually all magazines with annual incomes of over $2 million conducted research.
3. Among those magazines that conducted research, the median annual amount spent on research was about $8,000.
4. About one-third of the magazines who reported doing research have their own in-house research department; the remainder hire commercial firms to conduct their studies.

Folio's survey also revealed that the mail survey was the research method most often used by magazines; about 72% of the studies conducted by magazines that did research employed this method. Telephone interviews placed a distant second, accounting for only 7% of the studies, and the remaining 21% were performed using in-person interviews, panel studies, and focus groups. The type of test that was conducted most often was the market study. Figure 13.2 shows all the types of research projects conducted by magazines.

Summary

Magazine and newspaper research began in the early 1920s and for much of its early existence was qualitative in nature. Typical early research studies dealt with law, history, and international press comparisons. During the 1930s and 1940s, readership surveys and studies of the effectiveness of print media advertising were frequently done by private firms. By the 1950s, quantitative research techniques became common in print media research. The continuing competition with television and radio for advertisers and audiences during the last three decades has spurred the growth of private sector research. Professional associations have started their own research operations.

Research in the print media encompasses readership studies, circulation studies, studies of typography and makeup, and readability studies. Readership research is the most extensive area; it is conducted in order to determine who reads a publication, what items they read, and what gratifications they get from reading them. Circulation studies examine the penetration levels of newspapers and magazines in various markets as well as various aspects of the delivery and pricing systems. Typography and makeup are studied to determine the impact of different newspaper and magazine design elements on readership and item preferences. Readability studies investigate the textual elements within a message that affect comprehension of the message.

Figure 13.2 Types of research projects conducted by magazines

	Total	(Rank)
Market studies	57%	(1)
Reader profile/ demographic surveys	55	(2)
Surveys for editorial articles and features	54	(3)
Editorial effectiveness	44	(4)
Buying influence	36	(5)
Competitive publication analysis	34	(6)
Circulation profiling	32	(7)
Competitive readership studies	19	(8)
Product ownership and usage	17	(9)
Brand awareness/preference	15	(10)
Acquisition/launch analysis/appraisals	12	(11)
Advertiser/agency perception of media	11	(12)
Reader traffic studies (e.g., Starch, Readex)	5	(13)
Syndicated readership (e.g., SMRB)	2	(14)

Source: Barbara Love, "Folio Survey of Magazine Research," *Folio*, February 1981, pp. 54–61. Reprinted by permission.

Questions and Problems for Further Investigation

1. Assume you are the editor of an afternoon newspaper faced with a declining circulation. What types of research projects might you conduct to help you to increase your readership?
2. Now suppose you have decided to publish a new magazine about women's sports. What types of research would you conduct before starting to publish your magazine? Why?

3. Conduct a pilot uses and gratifications study of about 15–20 people to determine why they read the local daily newspaper.
4. Using this chapter as a sample, calculate the Flesch reading ease formula, the Gunning Fog Index, and McLaughlin's SMOG Grading.

References and Suggested Readings

Alperstein, G. 1980. "The Influence of Local Information on Daily Newspaper Household Penetration in Canada." *ANPA News Research Report* 26.

Bleyer, W. 1924. "Research Problems and Newspaper Analysis." *Journalism Bulletin* 1:1, 17–22.

Bogart, L. 1981. *Press and Public.* Hillsdale, N.J.: Lawrence Erlbaum and Associates.

Click, J. W., and Baird, R. 1979. *Magazine Editing and Production.* Dubuque: William C. Brown.

Dale, E., and Chall, J. S. 1948. "A Formula for Predicting Readability." *Education Research Journal* 27:1, 11–20.

Ferber, R. 1974. *Handbook of Marketing Research.* New York: McGraw-Hill.

Flesch, R. 1948. "A New Readability Yardstick." *Journal of Applied Psychology* 32:2, 221–33.

Fowler, G., and Smith, E. 1979. "Readability of Newspapers and Magazines over Time." *Newspaper Research Journal* 1:1, 3–8.

Gunning, R. 1952. *The Technique of Clear Writing.* New York: McGraw-Hill.

Haskins, J. 1960. "Pretesting Editorial Items and Ideas for Reader Interest." *Journalism Quarterly* 37:1, 224–30.

Haskins, J., and Flynne, L. 1974. "Effects of Headline Typeface Variation on Reader Interest." *Journalism Quarterly* 51:4, 677–82.

Hoskins, R. 1973. "A Readability Study of AP and UPI Wire Copy." *Journalism Quarterly* 50:2, 360–62.

Hvistendahl, J. K. 1977. "Self-Administered Readership Surveys: Whole Copy vs. Clipping Method." *Journalism Quarterly* 54:2, 350–56.

Larkin, E., and Hecht, T. 1979. "Research Assistance for the Non-Metro Newspaper, 1979." *Newspaper Research Journal*, Prototype Edition, 62–66.

Love, B. 1981. "*Folio* Survey of Magazine Research." *Folio*, February 1981, 54–61.

Lynn, J., and Bennett, E. 1980. "Newspaper Readership Patterns in Non-Metropolitan Communities." *Newspaper Research Journal* 1:4, 18–24.

McCombs, M., Mullins, L. E., and Weaver, D. 1974. "Why People Subscribe and Cancel: A Stop-Start Survey of Three Daily Newspapers." *American Newspaper Publishers News Research Bulletin* 3.

McCombs, M. 1977. "Newspaper Readership and Circulation." *ANPA News Research Report* 3.

McCombs, M. 1979. *Using Readership Research.* National Newspaper Foundation Community Journalism Textbook Project.

McLaughlin, H. 1969. "SMOG Grading: A New Readability Formula." *Journal of Reading* 22:4, 639–46.

Nafziger, R. 1930 "Reader-Interest Survey of Madison, Wisconsin." *Journalism Quarterly* 7:2, 128–41.

Nestvold, K. 1972. "Cloze Procedure Correlation with Perceived Readability." *Journalism Quarterly* 49:3, 592–94.

Nunn, C. 1979. "Newspapers with Separate Editorial Managements Have Higher Household Penetration." *ANPA News Research Report* 17.

Penrose, J., Weaver, D., Cole, R. and Shaw, D. 1974. "The Newspaper Non-Reader Ten Years Later." *Journalism Quarterly* 51:4, 631–39.

Poindexter, P. 1978. "Non-Readers, Why They Don't Read." *ANPA News Research Report* 9.

Poindexter, P. 1979. "Newspaper Readership and Circulation: An Update, 1977–79." *ANPA News Research Report* 22.

Schramm, W. 1957. "Twenty Years of Journalism Research." *Public Opinion Quarterly* 21:1, 91–108.

Schwartz, S., Moore, R., and Krekel, T. 1979. "Life Style and the Daily Paper: A Psychographic Profile of Midwestern Readers." *Newspaper Research Journal* 1:1, 9–18.

Siskind, T. 1979. "The Effect of Newspaper Design on Reader Preference." *Journalism Quarterly* 56:1, 54–62.

Sobal, J., and Jackson-Beeck, M. 1981. "Newspaper Nonreaders: A National Profile." *Journalism Quarterly* 58:1, 9–13.

Stone, G. 1978. "Literature Review: Using Community Characteristics to Predict Newspaper Circulation." *ANPA News Research Report* 14.

Stamm, K., Jackson, K., and Jacoubovitch, D. 1980. "Exploring New Options in Newspaper Readership Methods." *Newspaper Research Journal* 1:2, 63–74.

Swanson, C. 1955. "What They Read in 130 Daily Newspapers." *Journalism Quarterly* 32:3, 411–21.

Taylor, W. 1953. "Cloze Procedure: A New Tool for Measuring Readability." *Journalism Quarterly* 30:4, 415–33.

Tillinghast, W. 1981. "Declining Newspaper Readership: Impact of Region and Urbanization." *Journalism Quarterly* 58:1, 14–23.

Weaver, D., Wilhoit, C., and Reide, P. 1979. "Personal Needs and Media Use." *ANPA News Research Report* 21.

CHAPTER 14

Research in the Electronic Media

During the past decade research in the electronic media has expanded at a phenomenal rate. The 1982 *Broadcasting/Cable Yearbook* lists no fewer than 80 companies and individuals who are involved in some type of electronic media research. And this figure is not complete: hundreds of college and university professors and private citizens also conduct various types of studies connected with the electronic media. Add to this the in-house research conducted by stations and networks, and it is easy to see why broadcasting and cable research is now a multimillion-dollar business.

Electronic media research is changing continually, due to advancements in technology as well as improved research methodologies. The purpose of this chapter is to introduce some of the more widely used research procedures in this area.

Background

Although broadcasting is relatively young compared to other mass media, the amount and sophistication of broadcasting research have grown rapidly. During the initial years of broadcasting (the 1920s), there was little or no concern for audience research. The broadcasters were experimenters and hobbyists who were only interested in making sure that their signal was being sent and received. The potential popularity of radio was unknown, and there was no reason to be concerned with audience size at that time.

This situation changed rapidly during the 1930s as radio became a popular mass medium. When broadcast stations began to attract large audiences, there was a concern over how radio would be financed. Eventually it was decided that advertising was the most viable alternative (as opposed to government financing or taxes on sales of equipment). The acceptance of advertising on radio was the first step in the development of electronic media research.

Advertisers, not broadcasters, were the initiators of broadcast research. Once commercials began to be heard on the air, advertisers were naturally curious as to how many listeners were exposed to their messages and just how effective the messages were. Broadcasters were thus compelled to provide empirical evidence of their audience size and characteristics. This situation still exists—advertisers continually want more information about the people who hear and see their commercial announcements.

In addition to desiring information about the size of their audience, advertisers became interested in why people behave the way they do. This led to the development of the research area known as *psychographics*. But because they are rather vague, psychographics were not adequate predictors of audience behavior; advertisers wanted more information. Research procedures were designed to study *lifestyle* patterns and how they affect media use and buying behavior. Such information is valuable in designing advertising campaigns: if advertisers understand the lifestyle patterns of the people who purchase their products, then they can design commercials to match these lifestyles.

Electronic media research studies today fall into one of two main categories: *ratings* and *non-ratings* research. The remainder of this chapter is devoted to discussion of these two areas of research.

Ratings Research

When radio first became popular and advertisers began to see its potential for attracting customers, they were faced with the problem of documenting audience size. The print media were able to collect "hard" data, such as circulation figures, but broadcasters had no equivalent "hard" information—merely estimates. The early attempts at audience measurement failed to provide adequate data. Volunteer mail from listeners was the first source of data, but it is a well-known axiom of research that volunteers do not represent the general audience. Advertisers and broadcasters quickly realized that further information was urgently needed.

Since 1929, when the Crosley Radio Company conducted one of the first radio surveys, the bulk of ratings information for broadcast stations has come to be provided by two companies: A. C. Nielsen and The Arbitron Company. A third company, The Birch Report, has made great strides in broadcast ratings during the early 1980s and promises to offer formidable competition to Arbitron. Nielsen is the world's largest market research corporation, with offices in 23 countries; however, the ratings service accounts for only a small portion of its yearly sales of over $30 million. The Arbitron Company (ARB) is a subsidiary of the Control Data Corporation. The Birch Report, the product of former radio programmer Tom Birch, is based in Miami and involves only local market radio ratings.

The Nielsen Company collects several different types of ratings data for the broadcasting industry, including the National Television Index (NTI), the only *network* television ratings service in the country. Nielsen's National Station Index (NSI) provides local television market ratings, and the National Audience Composition (NAC) sample uses a diary approach to provide information about the people who watch television. These approaches are combined four times a year to produce **sweeps**, or surveys of every market in the country. Nielsen also provides **overnights** which are ratings surveys conducted in five cities: New York, Chicago, Philadelphia, Los Angeles, and San Francisco. Overnights are calculated using samples of about 100 households in each city. The preliminary data are computed via electronic meters so that network executives, program producers, and others may receive an indication of the performance of the night's programs very quickly; the final ratings appear several days later.

The Arbitron Company provides ratings for both local market radio and television stations; they are not concerned (at present) with network broadcasting. Arbitron also provides overnights in Chicago, Los Angeles, and New York. Although the Arbitron television service is widely used by stations and advertising agencies, the company is best known for its radio ratings. This trend may change after Arbitron expands its television services. At present, Arbitron is the primary company involved in local market radio ratings in the United States, but The Birch Report has made tremendous impact on the future of Arbitron. The only *network* radio ratings are gathered by Statistical Research, Inc., which is hired by networks to produce a RADAR report (Radio's All-Dimension Audience Research).

Broadcast ratings create controversy in many areas: viewers complain that "good" shows are continually being cancelled; producers, actors, and other artists complain that *numbers* are no judge of artistic quality (they are not intended to be); and advertisers often balk at the lack of reliable information. Although there may be merit to many of these complaints, one basic fact nevertheless remains: until further refinements are made, ratings as they currently exist will remain the primary decision-making tool in programming and advertising.

Since ratings will continue to be used for some time, it is important to understand several basic points about them. The first is that ratings are only estimates of audience size. They are not true indicators of anything. Too many people incorrectly attribute more importance to ratings figures than they deserve. All people who use ratings must understand that the numbers are only approximations; unfortunately, too many people consider them to be indicators of what actually exists (Head and Sterling 1982).

A second fact about ratings is that they are not always equally dependable: different companies produce different ratings figures for the same market during the same time period. Finally, ratings do not measure either the quality of programs or opinions about the programs. Ratings are estimates of audience size and nothing more.

The key point to remember when discussing or using ratings is that the figures are riddled with error. The data must be interpreted in light of several limitations (which are always printed in the last few pages of every ratings book). Individuals who use ratings as though they are "facts" are misusing the data.

Ratings Methodology

The research methodologies used by Nielsen, Arbitron, and Birch are complex; a detailed description of any one procedure would fill a small textbook. The data for ratings surveys are gathered by three basic methods: electronic recordings, diaries, and telephone interviews. Each method involves a slightly different sample selection procedure, and each has specific advantages and disadvantages.

Broadcast ratings provide a classic example of the need to sample the population. With just over 80 million households in the United States, it would be impossible for any ratings company to conduct a census of media use. The companies naturally resort to sampling to produce data which can be generalized to the population.

Nielsen's national sample for the NTI and NAC are selected using national census data and involve a procedure known as *multistage area probability sampling,* which insures that the sample reflects actual population distributions. That is, if Los Angeles accounts for 10% of the television households in the United States, Los Angeles households should compose 10% of the sample as well. Nielsen uses four stages in sampling: (1) selection of counties in the country; (2) selection of block groups within the counties; (3) selection of certain blocks within the groups; and (4) selection of individual households within the blocks. Twenty percent of the NTI sample of 1,170 households is reportedly replaced each year, as are one-third of the 3,200 households in the NAC sample.

For local measurements, both ARB and Nielsen use telephone directories as a sampling frame. The Nielsen samples are selected by computer from over 57 million residential telephones in the country. ARB also employs a systematic interval selection of telephone numbers for all households with television sets (see Chapter 4); however, because samples generated by telephone directories may not be representative of the population, ARB also uses an expanded sample frame (ESF), which selects samples from a list of randomly generated telephone numbers (Nielsen refers to this method as a "total telephone frame").

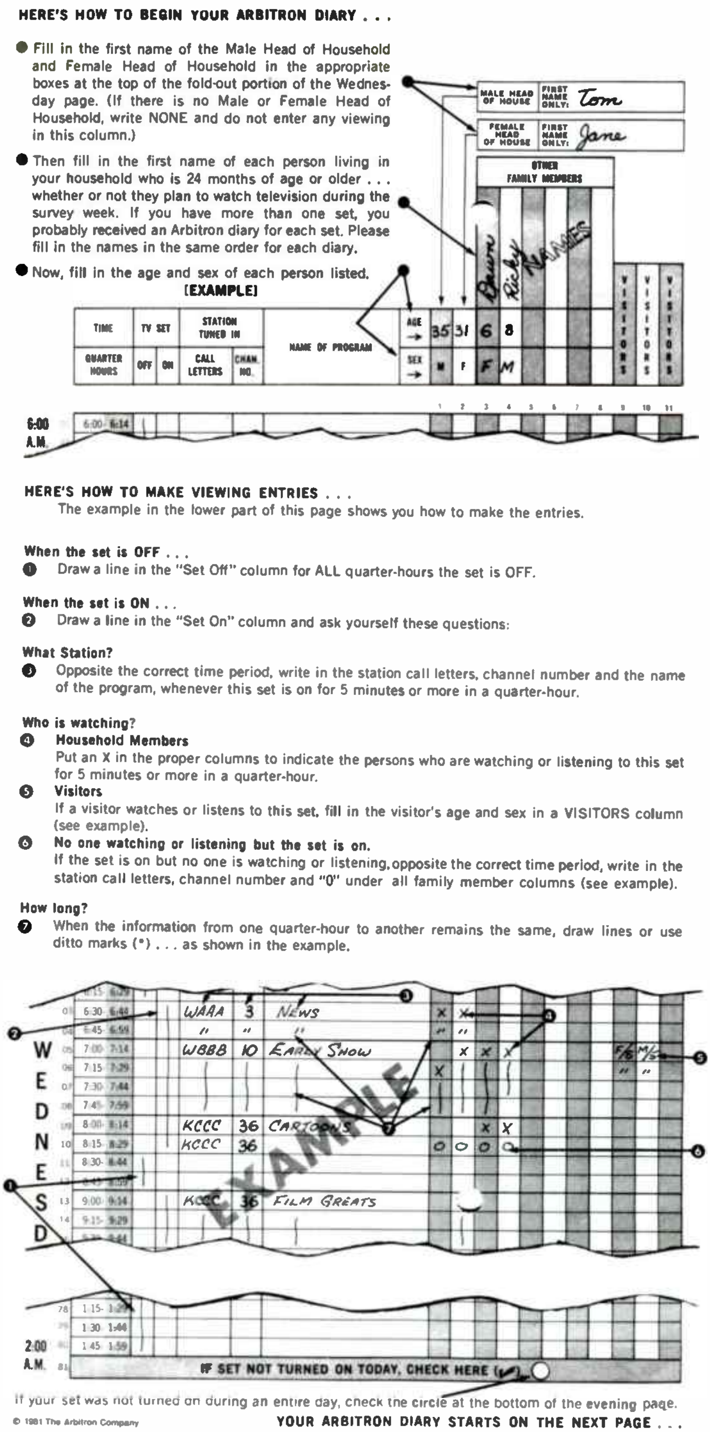
Target sample sizes for local audience measurements vary from market to market. ARB uses a formula to establish a minimum sample size, but there is no guarantee that this number of subjects will actually be produced. Many people may agree to participate in an audience survey, for example, but there is no way to force them to complete the diary they are given. Additionally, many completed diaries are often rejected because they are illegible or inaccurate. The companies are often lucky to get a 50% response rate in their local measurements.

In addition, each company obtains below-average response rates from minority groups. They consequently make an extra effort to collect data from these groups, contacting households by telephone or in person to assist them in completing the diary. (These methods are generally used in high-density Spanish and Black areas; otherwise, return rates could be too low to provide any type of audience estimates.) In such cases, *weighting* is also used; this will be discussed later.

Perhaps the most well-known method of gathering ratings data from a sample is by means of electronic ratings-gathering instruments, in particular the Nielsen *audimeter*, which was introduced in 1936 to record radio usage on a moving roll of paper. Today's audimeter, the Storage Instantaneous Audimeter (SIA), is a sophisticated device which automatically records the time and the broadcasting station, when each set in a household is turned on or off, how long the set stays on a channel, and all channel switchings. At least twice a day, each household in the NTI sample is called by the central computer, located in Dunedin, Florida, which retrieves the stored data and stores it for computation of the National Television Index. All of the data collection is done automatically and does not require participation by persons in the NTI households. The ratings results thus obtained from approximately 1,170 households are reported in trade journals and newspapers throughout the country. (The Arbitron Company also uses an electronic recording device for its television service, but uses a much smaller sample.)

The second major form of data collection is through the use of diaries, in which subjects are asked to record the channels they watch or the stations they listen to and for how long, as well as the number of people viewing or listening to each program. ARB uses diaries for both television and radio. Nielsen gives diaries to the 3,200

Figure 14.1 ARB ratings diary



HERE'S HOW TO BEGIN YOUR ARBITRON DIARY . . .

- Fill in the first name of the Male Head of Household and Female Head of Household in the appropriate boxes at the top of the fold-out portion of the Wednesday page. (If there is no Male or Female Head of Household, write NONE and do not enter any viewing in this column.)
- Then fill in the first name of each person living in your household who is 24 months of age or older . . . whether or not they plan to watch television during the survey week. If you have more than one set, you probably received an Arbitron diary for each set. Please fill in the names in the same order for each diary.
- Now, fill in the age and sex of each person listed.

HERE'S HOW TO MAKE VIEWING ENTRIES . . .

The example in the lower part of this page shows you how to make the entries.

When the set is OFF . . .

❶ Draw a line in the "Set Off" column for ALL quarter-hours the set is OFF.

When the set is ON . . .

❷ Draw a line in the "Set On" column and ask yourself these questions:

What Station?

❸ Opposite the correct time period, write in the station call letters, channel number and the name of the program, whenever this set is on for 5 minutes or more in a quarter-hour.

Who is watching?

❹ **Household Members**
Put an X in the proper columns to indicate the persons who are watching or listening to this set for 5 minutes or more in a quarter-hour.

❺ **Visitors**
If a visitor watches or listens to this set, fill in the visitor's age and sex in a VISITORS column (see example).

❻ **No one watching or listening but the set is on.**
If the set is on but no one is watching or listening, opposite the correct time period, write in the station call letters, channel number and "0" under all family member columns (see example).

How long?

❼ When the information from one quarter-hour to another remains the same, draw lines or use ditto marks (") . . . as shown in the example.

If your set was not turned on during an entire day, check the circle at the bottom of the evening page.

YOUR ARBITRON DIARY STARTS ON THE NEXT PAGE . . .

Leaving this portion of the page open will assist you in keeping the diary.

What TV stations (channels 2-83) can you receive clearly enough for viewing on this set? (Fill in below)

Channel Number	Station Call Letters	City	Channel Number	Station Call Letters	City

Accurate station identification is very important. Please use both the station call letters and channel number in your viewing entries.

MALE HEAD OF HOUSE	FIRST NAME ONLY:
FEMALE HEAD OF HOUSE	FIRST NAME ONLY:

OTHER FAMILY MEMBERS

NAMES

VISITORS VISITORS VISITORS

TIME	TV SET		STATION TUNED IN		NAME OF PROGRAM	AGE →											
QUARTER HOURS	OFF	ON	CALL LETTERS	CHAN NO		SEX →	M	F									
							1	2	3	4	5	6	7	8	9	10	11

3

6:00 A.M.

WEDNESDAY

	Quarter Hours	OFF	ON	CALL LETTERS	CHAN NO	NAME OF PROGRAM	1	2	3	4	5	6	7	8	9	10	11
01	6:00- 6:14																
02	6:15- 6:29																
03	6:30- 6:44																
04	6:45- 6:59																
05	7:00- 7:14																
06	7:15- 7:29																
07	7:30- 7:44																
08	7:45- 7:59																
09	8:00- 8:14																
10	8:15- 8:29																
11	8:30- 8:44																
12	8:45- 8:59																
13	9:00- 9:14																
14	9:15- 9:29																
15	9:30- 9:44																
16	9:45- 9:59																
17	10:00-10:14																
18	10:15-10:29																
19	10:30-10:44																
20	10:45-10:59																
21	11:00-11:14																
22	11:15-11:29																

3

DAYTIME

	Quarter Hours	OFF	ON	CALL LETTERS	CHAN NO	NAME OF PROGRAM	1	2	3	4	5	6	7	8	9	10	11
23	11:30-11:44																
24	11:45-11:59																
25	Noon-12:14																
26	12:15-12:29																
27	12:30-12:44																
28	12:45-12:59																
29	1:00- 1:14																
30	1:15- 1:29																
31	1:30- 1:44																
32	1:45- 1:59																
33	2:00- 2:14																
34	2:15- 2:29																
35	2:30- 2:44																
36	2:45- 2:59																
37	3:00- 3:14																
38	3:15- 3:29																
39	3:30- 3:44																
40	3:45- 3:59																

4:00 P.M.

Please be sure that all viewing has been recorded.

(Source: The Arbitron Company © 1981. Reprinted by permission.)

households in its NAC sample to supplement the information gathered from the SIA households, as the audimeter cannot record the number of people watching each set. An example of an ARB diary is shown in Figure 14.1.

Nielsen and Arbitron both use the telephone for custom research projects, where clients desire specific information quickly. However, **telephone coincidentals,** as these surveys are called, are considered to provide a very accurate assessment of audience viewing and listening data. Basically, the procedure involves selecting a number of households at random and calling these numbers during the viewing or listening period of interest. Individuals are simply asked what they are watching or listening to at that moment. This method avoids the problem of recalling information from the previous day (a method known as day after recall), or from several days prior to a survey period (Birch uses only telephone coincidentals).

Telephone coincidentals are sometimes tedious; calling households during prime time hours, for example, can be frustrating and nonproductive, since many people consider such calls an intrusion. For smaller research firms, however, telephone coincidentals offer an inexpensive, speedy, and accurate method of data collection.

Interpreting the Ratings

In order to explain the ratings interpretation process and its terminology, consider the following hypothetical analysis. (This example uses television networks, but the procedures are the same for radio ratings. In addition, the example has been simplified by using only three commercial television networks; local market ratings books will always include many more stations).

Assume that the Nielsen company collects these data for a certain *daypart* (time of day) on network television. Recall that Nielsen's NTI sample includes about 1,170 households in the United States and the data collected from them are generalized to the *total* population of more than 80 million households.

Network	*Households Viewing*
ABC	258
CBS	234
NBC	210
Not watching	468
	1,170

Rating

An *audience rating* is the percentage of people or households with television or radio in a population tuned to a specific station or network. Thus, the rating is expressed as the station or network's audience divided by the total number of television households in the population:

$$\frac{\text{People or Households}}{\text{Population}} = \text{Rating}$$

For example, ABC's rating using the hypothetical data is computed as:

$$\frac{258}{1{,}170} = .221 \text{ or } 22.1\%$$

This indicates that approximately 22.1% of the sample of 1,170 households was tuned to ABC at the time of the survey. (One point to note is that although ratings and related statistical values are percentages, the decimal points are eliminated when the data are reported to ease reading.)

The combined ratings of all the networks or stations during a specific time period provide an estimate of the total number of *homes using television* (HUT). Similarly, since radio ratings deal with persons rather than households, as in television, the term *persons using radio* (PUR) is used. The HUT or PUR can be found either by adding together the households or persons using either radio or television, or by computing the total rating and multiplying times the sample (or population when generalized). The total rating in the sample data is 59.9, which is computed as follows:

$$\text{ABC } \frac{258}{1{,}170} = .22 \text{ or } 22\%$$

$$\text{CBS } \frac{234}{1{,}170} = .20 \text{ or } 20\%$$

$$\text{NBC } \frac{210}{1{,}170} = .179 \text{ or } 17.9\%$$

HUT = 702   Total Rating = 59.9%

In other words, about 59.9% of all households with television were watching one of the three networks at the time of the survey. As mentioned, the HUT can also be computed by multiplying the total rating times the sample size: $.599 \times 1170 = 702$. This same formula is then used to project to the population. The population HUT is computed as: $.599 \times 80$ million $= 47.92$ million households.

Stations, networks, and advertisers naturally wish to know the estimated number of households in the HUT tuned to specific channels. The data from the sample of 1,170 households are again generalized to find households viewing for each network (or station):

Network	Rating	X	Population	=	Population HH Estimate
ABC	.22	X	80 million	=	17,600,000
CBS	.20	X	80 million	=	16,000,000
NBC	.179	X	80 million	=	14,320,000

Share

A share of the audience is the percentage of the HUT or PUR that is tuned to a specific channel or network. It is determined by dividing the number of households or persons tuned to a channel or network by the number of households or persons using their sets:

$$\frac{\text{People or Households}}{\text{HUT or PUR}} = \text{Share}$$

In the example, the sample HUT is 702 (258 + 234 + 210), or 59.9% of 1,170. The audience share for ABC would thus be:

$$\frac{258}{702} = .368 \text{ or } 36.8\%$$

That is, of the households in the sample whose television sets were turned on at the time of the survey, 36.8% were tuned to ABC (people may not actually have been *watching* the set in the case of electronic data gathering). The shares for CBS and NBC are computed in the same manner: CBS share = 234/702 or 33.3%; NBC share = 210/702 or 29.9%.

Shares are also used to estimate the number of households in the population viewing a station or network. The previous example demonstrating how to compute households is considered a "rough" estimate. However, there is often need for a more "exact" method for computing households. This is done by multiplying the sharc times the HUT or PUR. The exact households estimates for each network are computed as follows (rough estimates are for comparison):

Network	Share X	HUT	=HH in Population(exact)	(Rough Estimate)
ABC	36.8	X 47.92 m =	17,634,560	17,600,000
CBS	33.3	X 47.92 m =	15,957,360	16,000,000
NBC	29.9	X 47.92 m =	14,328,080	14,320,000
			47,920,000	47,920,000

Cost Per Thousand

Stations, networks, and advertisers need to be able to assess the efficiency of advertising on radio and television, to be able to determine which advertising "buy" is the most cost-effective. One common way to express advertising efficiency is in terms of *cost per thousand* (CPM), or what it costs an advertiser to reach 1,000 households or persons. The CPM provides no information about the effectiveness of a commercial message, only a dollar estimate of its reach. It is computed according to the following formula:

$$\text{CPM} = \frac{\text{Cost of Advertisement}}{\text{Audience Size (in thousands)}}$$

Using the hypothetical television survey, assume that a single 30-second commercial on ABC costs \$110,000. The CPM for such a commercial would be:

$$\text{ABC CPM} = \frac{\$110{,}000}{17{,}634.56\text{ M}} = \$6.23$$

Computing the CPM in the same manner for CBS and NBC, the figures are: CBS = \$6.89 CPM; NBC = \$7.67 CPM.

The CPM is regularly used when buying commercial time. Advertisers and stations or networks often negotiate an advertising contract using CPM figures; the advertiser might agree to pay \$7.00 per thousand households, for example. In some cases, no negotiation is involved; a station or network simply offers a program to advertisers at a specified CPM.

The CPM is generally not the only criterion used in purchasing commercial time, however. Other information, such as audience demographics and the type of program on which the advertisement will be aired, is also considered before a contract is signed. An advertiser may be willing to pay a higher CPM to a network or station that is reaching a more desirable audience for its product. Cost per thousand should be used as the sole purchasing criterion only when all else is equal: demographics, programming, advertising strategy, and so on.

Related Ratings Concepts

Although ratings and shares are important in audience research, there are a number of other computations that can be performed with the data. In addition, ratings, shares, and other figures are computed for a variety of survey areas and are split into several different demographic categories. For an additional fee, ratings companies will also provide "custom" information such as ratings according to ZIP code areas.

A **metro survey area** (MSA) generally corresponds to the Standard Metropolitan Statistical Areas (SMSA) for the country, as defined by the United States Office of Management and Budget. In other words, the MSA generally includes the town, county, or other designated area closest to the station's transmitter. The **area of dominant influence** (ADI), another area for which ratings data are gathered, defines each television or radio market in exclusive terms. Each county in the United States belongs to one and only one ADI, and rankings are determined by the number of television households in the ADI. Radio ratings use the ADIs established from television households; they are not computed separately. (Nielsen uses the term "designated market area," or DMA, to represent the ADI.)

The **total survey area** (TSA) includes the ADI and MSA as well as some other areas where the market's stations reach (known as *adjacent ADIs*). Broadcasters are most interested in TSA data because it represents the largest number of households or persons. In reality, however, advertising agencies look at ADI figures when purchasing commercial time; the TSA is infrequently used.

Ratings books are always divided into TSA, ADI, and MSA sections. Each area is important to stations and advertisers for various reasons, depending upon the type of product being advertised and the

Figure 14.2 Sample Arbitron ratings book

Daypart Audience Estimates Summary

Daypart and Station	TSA (000's) Children 2-11	6-11	Cumes (000's) TSA TV HH	WMN TOT	WMN 18-49	MEN TOT	ADI TV HH	ADI Ratings Persons TOT 2+	18+	18-49	12-24	12-34	18-34	Women TOT 18+	18-49	12-24	18-34	25-49	25-54	WKG WMN 18+	Men TOT 18+	18-49	18-34	25-49	25-54	TNS TOT 12-17	Child 2-11	6-11	Percent Distribution Metro	Home ADI	Adjacent ADI's #1	#2	#3	TV HH Rtgs in Adjacent ADI's #1	#2	#3
	31	32	111	118	119	128	105	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	62	65	66	67	68	70	71	72
MON-FRI 6.00P- 8 00P																																				
WCTV	14	8	159	162	96	136	113	26	30	23	17	20	22	31	25	19	24	27	28	29	29	22	20	24	26	14	18	18	38	82	9	2	7	5	2	6
WECA	2	1	31	30	16	25	25	3	3	2	2	2	2	3	2	2	2	2	3	3	3	2	1	2	3	3	3	3	40	82	18			1		
WVGA	1	1	12	10	9	8	4																				1		1	25	75			2		
WFSU			9	7	4	5	7																						34	71			29			
WALB							42	3	4	3	2	2	2	4	3	2	2	3	4	5	4	2	2	3	4	2	1	1								
* WMBB							22	2	2	2	2	2	2	2	2	2	2	2	2	1	1	2	1	2	2	2	2	3								
WTBS							19	2	2	2	3	3	2	2	2	3	2	2	2	2	2	2	3	2	2	3	3	4								
TOT/PVT	17	10	359	375	231	326	128	40	44	34	27	32	33	46	36	30	36	40	43	42	42	32	30	36	40	27	31	32	37	80	11	2	6			
7 00P- 7 30P																																				
WCTV	21	13	106	106	61	86	81	28	29	23	21	22	22	31	26	26	26	26	27	30	27	20	19	22	23	23	27	27	34	78	13	3	6	7	2	5
WECA	2	2	17	17	10	13	13	3	3	2	2	2	2	4	3	2	3	3	4	4	3	1	1	1	3	3	3	4	48	78	22			2		
WVGA			5	4	3	3	1																						4	14	86			1		
WFSU			5	4	2	3	4						1	1			1	1	1	1									34	61			39			1
WALB							23	5	6	4	2	3	3	7	4	2	3	5	6	7	6	4	3	6	7	1	1									
WMBB							5	1	1	1		1	1	1	1		1	1	1							1	1	1								
WTBS							12	3	3	3	3	3	3	3	3	3	3	3	3	2	3	3	4	3	4	3	4	5								
TOT/PVT	23	15	332	330	197	279	117	43	45	35	29	33	34	47	38	34	39	41	44	43	43	32	30	37	42	31	38	40	35	77	14	2	6			
7 30P- 8 00P																																				
WCTV	18	12	109	93	57	83	81	22	24	21	17	20	22	24	22	17	23	23	23	25	24	21	22	23	23	13	20	23	38	77	16	1	6	8	1	5
WECA	5	2	21	22	11	16	17	6	6	4	5	5	4	6	4	7	5	5	6	7	6	4	3	4	6	7	7	6	43	85	15			2		
WVGA	2	1	7	5	4	6	3	1	1	1	1	1	1	1	1	2	1	1	1	1		1	1			2	2	1	1	53	47			1		
WFSU			5	3	1	3	3																						44	68			32			
WALB							32	5	5	2	2	2	2	5	3	3	2	3	4	5	4	2	2	3	3	4	4	3								
WMBB							14	3	3	3	3	3	2	3	3	3	3	4	4	2	3	3	2	3	4	5	4	5								
WTBS							8	2	1	2	2	2	2	1	2	2	1	2	2	1	2	2	2	2	2	4	4	5								
TOT/PVT	25	15	340	331	199	289	122	45	45	37	34	39	38	46	38	35	40	42	45	45	45	36	36	40	44	39	47	49	37	77	17	1	5			
8 00P- 11 00P																																				
WCTV	17	10	181	187	118	155	117	23	24	23	21	23	24	26	26	24	28	27	27	28	21	20	20	22	22	20	20	21	35	76	18		5	9		4
WECA	2	1	52	55	38	47	48	5	6	6	6	6	7	6	7	6	7	6	6	7	6	6	6	6	6	4	3	3	56	93	6		1	1		
WVGA	1		19	19	16	16	6	1	1	1	1	1	1	1	1	1	1				1	1	1			1	1	1	1	36	64			2		
WFSU			18	18	9	11	13							1			1	1	1										28	56		4	40			1
WALB							63	8	8	5	5	5	5	7	6	6	6	5	6	6	6	4	4	5	6	6	4	4								
WMBB							32	3	3	3	2	3	3	3	3	2	3	4	4	3	3	3	3	4	3	2	3	3								
WTBS							12	1	1	1	1	1	1					1			1	1	1	1	1			1								
TOT/PVT	20	11	362	385	241	335	128	43	45	43	38	43	45	48	47	42	50	47	48	49	42	39	39	44	46	38	33	34	37	76	18		5			
11 00P- 11 30P																																				
WCTV	2	1	77	79	52	57	60	12	15	14	12	13	14	17	16	14	18	16	17	19	13	11	11	11	10	7	2	2	49	84	10		6	3		3
WECA			10	10	7	8	10	1	1	1	1	1	1	2	2	1	1	2	2	1	1	1		1	2				71	99						
WVGA			4	3	2	3	3				1	1				1										1				81	19					
WFSU			2	2	1	1	2																						42	52			48			1
WALB							12	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2											
WMBB							4																													
WTBS							1																													
TOT/PVT	2	1	228	217	133	178	76	17	21	19	15	17	19	23	21	17	23	22	23	22	18	16	16	17	16	10	2	2	49	84	9		6			
11 30P- 1 00A																																				
WCTV	1		41	43	28	29	31	3	4	4	5	4	5	5	5	6	6	5	5	5	3	4	3	3	2	3	1	1	48	83	12		5	1		1
* WECA			6	4	3	4	6				1	1	1					1					1			1			52	98	2					
WVGA			4	3	2	4	3																1					1	11	66	34					
* WFSU			1	1	1		1																						61	61			39			
* WALB							8	1	1	1	1	1	1	1							1	1	1	1												
WMBB							4																	1												
WTBS							2																													
TOT/PVT	1		157	142	85	118	47	5	6	7	7	7	7	7	7	7	8	6	7	6	6	6	7	5	4	4	2	2	47	82	12		6			
SATURDAY 8 00A- 1 00P																																				
WCTV	22	13	49	17	12	17	29	7	3	3	6	5	3	3	4	8	4	3	3	3	2	3	3	3	3	10	25	26	35	63	31		6	6		2
WECA	3	2	10	3	2	5	8	1	1	1	1	1	1	1	1	2	1	1		1		1	1	1	1	3	4	4	47	90	10					
* WVGA			4	2	2	1	1																							18	82			1		
WFSU			4	3	1	1	3																						54	74			27			
WALB							11	1		1	2	1	1			1					1	1	1			3	5	5								
WMBB							3																				1	1								
WTBS							5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1											
TOT/PVT	25	15	163	81	54	66	48	13	6	6	12	10	7	6	7	14	8	5	5	5	5	6	6	5	5	18	39	41	35	64	29		6			
SAT+SUN 1 00P- 5 00P																																				
WCTV	5	3	73	56	36	53	43	5	6	4	5	5	5	6	5	6	6	4	5	7	5	4	4	4	4	4	6	6	25	57	37	1	4	7		1
WECA	1		20	8	4	17	19	2	2	1	2	2	1	1	1	1	1	1	1	1	2	2	2	3	3	3	1	1	53	98	2					
* WVGA			6	4	3	3	2																							32	68					
WFSU			7	8	4	3	5												1	1									34	60			40			1
* WALB							13	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1											
WMBB							5		1												1	1	1		1											
WTBS							10	1	1	2	2	1	2	1	1	1	1	1	1	1	2	2	2	1	1	1	1	1								
TOT/PVT	6	3	219	175	104	173	73	12	13	11	13	12	12	11	10	12	10	9	10	12	15	13	13	12	13	13	9	9	29	63	31	1	5			

ADJACENT ADI # 1 - ALBANY, GA
ADJACENT ADI # 2 - DOTHAN
ADJACENT ADI # 3 - PANAMA CITY

Daypart Summary

(Source: The Arbitron Company. Reprinted by permission.)

goals of the advertising campaign. A station may find, for instance, that it is very strong in the MSA but weak in the ADI. Knowing this information, the station could plan promotional events and other activities to help increase its ratings outside the metro survey area.

An example of an Arbitron television ratings book is shown in Figure 14.2. Reading a ratings book is relatively simple. As mentioned earlier, all decimal points have been deleted. The upper right hand portion of the table indicates that the numbers are from a "Daypart Summary" section of the book. A **daypart** is defined as a segment of a broadcast day, such as "prime time" (8:00–11:00 EST).

In order to read the numbers, one first looks at the top of each column to find out what the numbers represent. For example, Column 31 represents the number of children between the ages of 2 and 11 estimated to be in the audience for each station. Columns 35–40 display the ADI ratings for all viewers; in Columns 41–52, the ratings are broken down for men and women at different age levels. In every ratings book, all terms are identified in a glossary included at the end of the book. Some of the more widely used terms are discussed below.

The **average quarter hour** (AQH) is an estimate of the number of persons or households tuned to a specific channel for at least five minutes during a 15-minute time segment. These estimates are provided for the TSA, ADI, and MSA in all ratings books. Stations are obviously interested in obtaining high AQH figures in all demographic areas, since these figures indicate how long an audience is tuned in, and thus how "loyal" the audience is to the station.

The **cume** (cumulative audience) is an estimate of the number of different persons who listened to or viewed at least five minutes within a given daypart. The cume is also referred to as the "unduplicated audience." For example, a person who watches a soap opera at least five minutes each day Monday through Friday would be counted only once in a cume rating, whereas his or her viewing would be "duplicated" five times in determining average quarter hours.

The **gross rating point** (GRP) is a total of a station's ratings during two or more dayparts and represents the size of the gross audience. Advertising purchases are often made on the basis of GRPs. For example, a radio advertiser who purchases ten commercials on a station may wish to know the gross audience they will reach. Using hypothetical data, the GRP is calculated as:

Daypart	No. of Spots		Station Rating		GRP
M–F, 6–9a.m.	2	×	3.1	=	6.2%
M–F, 12–3p.m.	2	×	2.0	=	5.8%
M–F, 1–6p.m.	2	×	3.6	=	7.2%
Sat, 6–9a.m.	2	×	2.5	=	5.0%
Sun, 3–6p.m.	2	×	4.1	=	8.2%
	10				32.4%

The gross rating point indicates that about 32.4% of the listening audience will be exposed to the ten commercials.

Audience Turnover. A useful figure for radio stations is the **audience turnover,** or the number of times the audience changes during a given daypart. A high turnover is not always a negative factor in advertising sales; some stations naturally have high turnover (such as "Top 40" stations, whose audiences comprise mainly younger people who tend to listen for short periods of time before they change channels). A high turnover factor simply means that an advertiser will need to run more spots to reach the station's audience. Usually such stations will compensate for this by charging less for commercial spots than stations with low turnovers.

Turnover is computed by dividing a station's cume audience by its average persons total (both of these figures are reported in ratings books). Consider the following example of three stations in the Monday–Friday, 3:00–6:00 p.m. daypart.

Station	Cume Audience		Average Persons		Turnover
A	2,900	÷	850	=	3.4
B	1,750	÷	420	=	4.2
C	960	÷	190	=	5.1

In this market, an advertiser on Station C would need to run more commercials to reach all listeners than one who uses Station A. However, Station C, in addition to having a larger audience, may have the demographic audience most suitable for the advertiser's product.

Adjusting for Unrepresentative Samples

Since ratings are computed using samples from the population, there is always a certain amount of error associated with the data. This error, designated by the notation σ, is known as standard error (introduced in Chapter 4). Standard error must always be considered prior to interpreting ratings in order to determine whether a certain sex/age group has been undersampled or oversampled.

There are numerous approaches to calculating standard error. One of the simpler methods is:

$$SE(p) = \sqrt{\frac{p(100 - p)}{n}}$$

where p = sample percentage or rating

n = sample size

SE = Standard Error

For example, suppose a random sample of 1,200 households produces a rating of 20. The standard error can be expressed as:

$$SE(p) = \sqrt{\frac{20(100 - 20)}{1,200}}$$

$$= \sqrt{\frac{20(80)}{1,200}}$$

$$= \sqrt{1.33}$$

$$= \pm 1.15$$

The rating of 20 has a standard error of ±1.15 points, which means the rating actually ranges from 18.85 to 21.15. Standard error formulas are included in all ratings books; Arbitron has simplified the procedure through the use of tables in the back of each book.

Weighting is another procedure used by ratings companies to adjust for samples which are not representative of the population. In some situations a particular sex/age group cannot be adequately sampled, and it becomes imperative that a correction be made.

Assume that population estimates for an ADI indicate that there are 41,500 men aged 18–34, and that this group accounts for 8.3% of the population over the age of 12. The researchers distribute diaries to a sample of the ADI population of which 950 are returned and usable (known as *in-tab* diaries). They would expect about 79 of these to be from men aged 18–34 (8.3% of 950). However, they find that only 63 of the diaries are from this demographic group—16 short of the anticipated number. The data must be weighted to adjust for this deficiency. The weighting formula is:

$$\text{Weight}_{(\text{sex/group})} = \frac{\text{Group Share of Population}}{\text{Group Share of Sample}}$$

In the example, the weighting for men aged 18–34 is calculated as:

$$\text{Weight}_{(\text{MSA Men } 18-34)} = \frac{.083}{.066}$$

$$= 1.25$$

To interpret this figure, it must be multiplied by the number of persons in the group that each diary would normally represent. That is, instead of representing 525 men (41,500 ÷ 79), each diary would represent 656 men (525 × 1.25). The ideal weighting value is 1.00, indicating that the group was adequately represented in the sample. On occasion, a group may be oversampled, in which case the weighting value will be a number less than 1.00.

Recent Developments in Ratings Research

A newcomer to television ratings research is Information Resources, Inc., which has developed a system known as BehaviorScan. This system is a two-way interactive methodology whereby television signals are sent via cable to respondents' households and the respondents phone in data. The company not only collects ratings information but also correlates the respondents' purchasing patterns and behavior with the television programs they watch (that is, with the commercials to which they are exposed during these programs). All evidence indicates that the two-way interactive system will be the basis for television ratings in the future.

The rapid growth of cable television and the audience fragmentation it has produced have created a desire by advertisers and cable operators to have more ratings information than Arbitron and Nielsen provide. Multiple channel cable systems offer great flexibility for viewers and programmers, but they also create difficulty for ratings companies in trying to estimate audience size and composition. In 1981, the Home Testing Institute initiated the Video Probe Index, which is designed solely for cable television networks and includes estimates of daily, weekly, and monthly cume ratings three times a year. The index utilizes a national nonrandom sample of 6,000 households that answer questionnaires mailed to them; the methodology is "channel recall," a procedure similar to that used in magazine advertising recall (*Advertising Age,* December 14, 1981).

Non-Ratings Research

Although audience ratings are the most visible research data used in broadcasting, there are numerous other methodologies used by broadcasters, production companies, advertisers, and broadcast consultants. Ratings provide estimates of audience size and composition. Non-ratings research provides information about what the audience likes and dislikes, analyses of different types of programming, demographic and lifestyle information about the audience, and much more. All of these data are intended to provide decision-makers in the industry with information they can use to eliminate some of the guesswork.

Non-ratings research cannot solve all the problems broadcasters face, but it can be used to support decision-making. This section describes some of the non-ratings research that is conducted in the electronic media.

Program Testing

Research has become an accepted step in the development and production of programs and commercials. It is now common practice to test these productions in each of the following stages: (1) initial idea or plan; (2) rough cut; and (3) postproduction. A variety of research approaches can be used in each of these stages, depending upon the purpose of the study, how much time is allowed for testing, and what types of decisions will be made with the results. One of the responsibilities of a research director is to determine what the decision-makers need to know and then design an analysis to provide that information.

Since the production of any major program or commercial is a very expensive procedure, producers and directors are interested in gathering preliminary reactions to a planned project. It would be ludicrous to spend thousands or millions of dollars on a project which would have little audience appeal. One way to collect preliminary data is to show subjects a short statement summarizing a program or commercial and then question them as to their opinions about the idea and their willingness to watch the program or buy the product on the basis of the description. The results may provide some indication about the potential success of a show or commercial.

The problem with program descriptions, however, is that they cannot demonstrate the characters in the program who may make the show a success or a failure. For example, the ABC program "Mork and Mindy" might have been described as:

> *Mork and Mindy:* The hilarious comedy series about the relationship between Mindy, a single attractive woman who lives in a Boulder apartment, and Mork, a rather zany and unpredictable young man who came from the planet Ork. Mork landed on earth (in a space ship resembling a large chicken egg) to investigate human behavior and report back to his superiors. Each week the program revolves around Mork's encounters with various aspects of human behavior and portrays how he learns about the American way of life.

To many people this statement might seem to describe the type of show generally referred to as a "bomb." However, the indescribable on-the-air relationship and dialogues between the two main characters made "Mork and Mindy" a hit show in the late 1970s. If producers relied totally upon program descriptions in testing situations, many successful shows would never have reached the air.

If an idea tests well in the preliminary stages (or if the producer or advertiser wishes to go ahead with the project regardless of what the research indicates), the next step is to produce a model or simulation of the product. These are referred to as *rough cuts, storyboards, photomatics, animatics,* or *executions.* The **rough cut** is a simplistic production that usually uses amateur actors, little or no editing, and makeshift sets. The other models are photographs, pictures, or drawings of major scenes designed to give the basic idea of a program or commercial to anyone who looks at them.

Rough cuts or models are tested by companies such as Burke Marketing Research. They do not involve a great deal of production expense, which is especially important if the tests show a lack of acceptance or understanding of the product. The tests provide information about the script, characterizations, character relationships, settings, cinematic approach, and overall appeal. In most cases they do not identify the causes when a program or commercial is found to be unacceptable to the test audience; rather, they merely provide an overall indication that something is wrong.

Once the final product is available, postproduction research can be conducted to determine how well it is received by the audience. Finished products are tested in experimental theaters; in shopping centers, where mobile vans are used to show commercials or programs; at subjects' homes in cities where cable systems provide test channels; or via telephone, in the case of radio commercials.

Results from postproduction research often indicate that, for example, the ending of a program is unacceptable and must be reedited or reshot. Many types of problems that were not foreseen during production may be encountered in postproduction research, and the data usually provide producers with an initial audience reaction to the finished, or partially finished, product.

One of the more well-known experimental theaters used for testing television programs and commercials and theatrical films is run by Audience Studies Incorporated (ASI): the ASI Preview House in Hollywood, California. Preview House has a capacity of 400, and each seat contains a dial with which the viewer rates the material he or she watches. The responses, from "very good" to "very dull," are recorded by a computer in the back of the theater. The data are then displayed in a graph form that enables researchers to determine if the material contains any problem areas when compared to similar material that has been previously tested. Preview House always begins each testing session with a "Mr. Magoo" cartoon that serves as a baseline for comparison for further testing and also allows the audience to learn to operate their dial. For this reason, the results of these tests are referred to as "Magoos." Although Preview House is often criticized for its methodology, sample selection, and quality of data, most broadcast and film executives seem to think that its research provides at least some indication of audience likes and dislikes.

Another postproduction research company is Vox Box, located in Seattle, Washington. Vox Box uses a two-way cable system that allows households in the sample of 500 to respond to commercials, pilot programs, and finished productions by pressing buttons on a hand-held device similar to a calculator. The Vox Box system provides producers and advertisers with data that are collected immediately from audiences viewing programs in their own homes.

Program testing is constantly being conducted by a number of different companies. The results are used to schedule programs and commercials, change characters in (or add new ones to) established programs, rewrite scripts, and keep up to date with current fads and trends. (An example of the latter is the Yankelovich Monitor, a syndicated report of social trends in the United States.)

Although most program testing is conducted by major networks, large advertising agencies, and production companies, there is an increasing interest in this area of research at the local level. Stations now test promotional campaigns, prime time access scheduling, the acceptability of locally produced commercials, and various programming strategies.

Music Call-out Research

Music is obviously one of the most important elements in a radio station's programming schedule. In order to provide listeners with music they like to hear and to eliminate songs that people are tired of hearing (known as burn-out), radio programmers engage in a variety of research procedures. One of the most widely used methods is **call-out research,** whereby listeners are telephoned and asked their opinions about certain music selections.

The first step in call-out research is to develop a random sample of listeners (or non-listeners). The best way to do this is by using some form of random digit dialing. However, many programmers use sample frames composed of contest winners, people who have called in to request music, and so forth. This type of list (as discussed in Chapter 4) does not give all listeners an equal chance of being selected and should be used with caution.

Once the sample is selected, a recording is made of the music to be tested. Only a short segment, or **hook,** of each recording is used. A hook is a 5–15-second representative sample of the recording—enough for a person to determine whether he or she would like to hear the song played more often, or is tired of hearing it. Subjects who agree to participate are told they will hear several musical selections over the telephone which they will be asked to rate according to a predetermined scale.

Several types of measurement scales are appropriate for use in call-out research. For example, the respondent can be asked to rate a selection on a five- or seven-point scale where "1" represents "I did not like the selection" and "5" or "7" represents "I liked the selection a lot." A feeling thermometer might also be used to rate selections, with "0" indicating an unfavorable response and "100" a favorable response. Scores for each musical selection are tabulated to determine which ones appealed most to listeners. These data are used by program directors as a guide in developing their station's playlist.

Call-out research is a valuable research tool for programmers: data can be collected on the attitudes listeners have toward new or old music, performers or groups, or unique musical tastes. Also, since the research is generally conducted by the station's staff, reliable data can be collected with a minimum investment of time and money.

News and Programming Consultants

There are several private consulting firms that conduct news and programming research for radio and television stations. Some of the largest are Frank N. Magid Associates, McHugh and Hoffman, Reymer and Gersin, and the Simmons Market Research Bureau. Although each firm specializes in certain areas of broadcasting and uses different research procedures, all have a common goal: to provide broadcast management with data to be used in decision-making processes.

Consultants provide custom research for broadcasters; they perform studies of news formats and their popularity, news personalities, programming formats in radio, and program scheduling in both radio and television, as well as individual research studies developed by management. Data are collected in a variety of ways. Usually a company will settle on one or two basic methods such as telephone interviews, focus groups, in-person interviews, or intercepts.

Custom research is not conducted solely by large companies; there are a number of highly qualified individuals who also provide research to broadcasters. Usually these consultants have worked in the industry for several years before they branch out into research. There are also a number of college and university professors who conduct research on a variety of topics.

Performer Q

Producers and directors in broadcasting naturally desire an indication of the popularity of various performers and entertainers. A basic question in the planning stage of any program is, "What performer or group of performers should be used to give the show the greatest appeal?" It is not unreasonable to expect producers to prefer using the most popular and likeable performers in the industry to taking a chance on an unknown entertainer.

Market Evaluations, Inc. meets the demand for information about performers and entertainers. The company conducts nationwide surveys of about 1,000 families to develop two types of ratings: the familiarity rating and the likeability rating (the latter more widely known as the "Performer Q"). The ratings comprise 17 different categories of entertainment and provide information about how the general public perceives various performers and artists: the higher the "Q" rating, the more popular and likeable the performer.

Focus Groups

The focus group, discussed in Chapter 6, is becoming increasingly popular in electronic media research, probably due to its versatility. Focus groups are used to develop questionnaires for further research and to provide preliminary information on a variety of topics, such as format and programming changes, personalities, station images, and lifestyle characteristics of the audience. The latter is particularly useful when the focus group consists of a specific demographic group.

Miscellaneous Research

Broadcast stations are unique and require different types of research. Some of the kinds of research conducted by and for stations include:

1. *Ascertainment research.* Although the Federal Communications Commission has begun to change the status of ascertainment, many radio stations and all TV stations continue to conduct ascertainment studies. In most situations the telephone is used to collect data for general community ascertainment; personal interviews are conducted for the community leader portion. In order to make full use of ascertainment research, most stations add questions to their survey to collect additional information on programming likes and dislikes, station image, and other areas of interest.

2. *Station image.* Station image has been mentioned in several places throughout this chapter. It is important for a station's management to know how the public perceives the station and its services. Public misperception of management's purpose can create a decrease in audience size and consequently in advertising revenue. For example, suppose a radio station which has aired "Top 40" music for ten years decides to switch to a country format. It is important that the audience and advertisers be aware of this change and have a chance to voice their opinions. This can be accomplished through a station image study, where respondents to telephone calls are asked questions such as: "What type of music does radio station WAAA play?" "What types of people do you think listen to WAAA radio?" and "Did you know that WAAA now plays country music?" If the research reveals that few people are aware of the change in format, management can develop a new promotion strategy. Or the station might find that the current promotional efforts have been successful and should not be changed.

Station image studies are conducted periodically by most larger stations in order to maintain current information on how the audience perceives each station in the market. It is important for station managers to keep up to date with audience trends and social changes in order to provide the services that listeners and viewers want.

3. *Advertiser (account) analysis.* In order to provide advertisers with the best possible service, many stations administer questionnaires which seek information about local businesses. Some typical questions include: "When did your business open?" "How many people are involved in ownership of this business?" "How much do you invest in advertising per year?" "When are advertising purchase decisions made?" and "What do you expect from your advertising?"

Information obtained from client questionnaires is used to help write more effective advertising copy, to develop better advertising proposals, and to allow the sales staff to know more about each client. Generally, the questionnaires are administered before an advertiser first uses the station, but they can also be used for advertisers who have done business with the station for several years.

4. *Sales research.* In an effort to increase sales, many stations take it upon themselves to conduct research for local clients. For example, a station may conduct a "banking image" study of all banks in the area to determine how residents perceive each bank and the service it provides. The results from such a study are then used in an advertising proposal for the banks in the area. If a station discovers that the 1st National Bank's 24-hour automatic teller service was not well understood by local residents, the station could develop an advertising proposal to concentrate on this point.

5. *Diversification analyses.* The goals of any business are to expand and to achieve higher profits than it did in the previous year. In an effort to reach these goals, most larger stations, partnerships, and companies engage in a variety of studies to determine where investments should be made. Should other stations be purchased? What other types of activity should the business invest in? Such studies are used for forecasting and represent a major portion of the research undertaken by larger stations and companies.

The Future of Broadcast Research

Trying to predict the future of broadcast research is no simple task. Certain trends are clear, but long-range predictions are tenuous because the media are changing so rapidly.

It is no secret that the position of radio and television in American society is changing. In an effort to maintain audience interest, radio stations are reconsidering their significance in the broadcasting world. AM radio stations are pushing for stereo transmission to increase their marketability, while FM stations are trying to develop new alternatives in order to maintain their stronghold on the medium. Some critics claim that radio is "dead," but history has proven that neither AM nor FM radio managers are likely to stand by and let other services take over. Radio broadcasters for the most part know that change is necessary to survive, and change is what they are investigating through various research projects designed to answer such questions as: "How can we more effectively position the station to attract a solid and identifiable audience?" "How can other delivery systems, such as cable radio, help to increase our reach?" and "Can satellite transmission be used effectively?" Additionally, they are searching for more detailed and refined audience measurement techniques to provide advertisers with higher quality data and better information about programming: what music to play, how long selections should be kept on the air, and how often recordings should be rotated to maintain audience interest.

Television offers an even more confusing picture than radio. The decade of the 1970s provided the stepping stone to the future of television as new technologies paved the way for what was and is to come: direct broadcast satellite transmission; multitiered cable television operations; video games; videotape recorders; videodiscs; television component systems; multiple-use television sets, which serve as security systems, telephone extensions, and so forth; and multiple-screen sets, which allow viewing of several channels simultaneously.

All of these developments lead to the same conclusion about the future of research in the electronic media: there will be a need for information about how people are using all of these new technologies. How is conventional television use changing? What effects do all of these new devices have on consumers? Some of the specific questions that will need research include:

1. What effect do cable systems that offer 50+ channels have on consumers' utilization of television? What do they watch? Why do they select specific channels?
2. What kinds of programming will fill these 50+ channels? Can cable companies afford all of these channels? What is the dollar limit that subscribers will pay for premium (or basic) services?
3. To what degree are audiences segmented by the new technologies? Will advertisers be willing to sponsor programs aimed at a narrowly defined audience? What will happen to commercial network advertising when such a variety of choices becomes available?
4. How do videotape and videodisc households differ in their use of television from homes without such devices?
5. How much will people be willing to spend for their own satellite dish?
6. What are the limits of television use that the average consumer can handle? Is there a "breaking point" at which viewers will just turn off their sets?
7. Will superstations such as WTBS continue to be successful?
8. Will there be several sports channels, news channels, or weather channels? What effects will these services have on consumer use of television?

The list of questions could continue for several pages. The main point is that there are many things that need investigation. The electronic media pose a literal gold mine of questions for researchers to help answer.

In addition to the variety of questions that need investigation, researchers will need to develop new methodologies. Although the same procedures used today will be adequate for many of the questions of the future, the need for new research methods is clear. How will audience measurements be conducted when viewers (and listeners) can select from so many channels? What methods are appropriate to help advertisers make better judgments in purchasing advertising slots? What type of media mix in advertising is the most efficient? Again, the list is almost endless. But the future of electronic media research seems clear: new methodologies will be required to answer new research questions.

Summary

The purpose of this chapter has been to introduce some of the more common methodologies used in broadcast research. Ratings are the most visible research used in broadcasting as well as the most influential in the decision-making process. However, other forms of research such as focus groups, call-out research, image studies, and program testing are all used frequently in order to collect data. The importance of research is fueled by an ever-increasing desire by management to learn more about broadcast audiences and their uses of the media.

The phenomenon of audience fragmentation has become evident during the past several years. Each medium now attracts a smaller, much more narrowly defined group of people. This competition for viewers and listeners has created a need for research data. Broadcast owners and managers realize now that they can no longer rely on "gut feelings" when making programming, sales, and marketing decisions. The discussions in this chapter have been designed to emphasize the importance of research in all areas of broadcasting.

Questions and Problems for Further Investigation

1. Assume a local television market has three stations: Channel 2, Channel 7, and Channel 9. There are 200,000 television households in the market. A ratings company samples 1,200 households at random and finds that 25% of the sample is watching Channel 2, 15% Channel 7, and 10% Channel 9.
    a. Calculate each station's share of the audience.
    b. Project the total number of households in the population watching each channel.
    c. Calculate the CPM for a $1,000, 30-second spot on Channel 2.
    d. Calculate the standard error involved in Channel 2's rating.
2. What are the major data-gathering problems associated with: (a) electronic recorders; (b) diaries; and (c) telephone coincidentals?
3. Examine a recent ARB market radio ratings book. Select a station and a daypart and compute that station's: (a) AQH; (b) cume; and (c) turnover.
4. Perform your own music call-out research. Edit several 15-second selections of new recordings on a reel or cassette and ask people to rate them on a seven-point scale. Compute means and standard deviations for the results. What can you conclude?
5. Several questions relevant to modern broadcasting trends are listed at the end of the chapter. What type or types of research methods could be used to try to answer these questions?

References and Suggested Readings

Arbitron. 1980. *Description of Methodology.* The Arbitron Company.

Broadcast Ratings Council. 1975. *Maintaining Rating Confidence and Credibility.* New York: Broadcast Ratings Council.

Head, S., and Sterling, C. 1982. *Broadcasting in America.* 4th ed. Boston: Houghton Mifflin.

Fletcher, J. E., ed. 1981. *Handbook of Radio and TV Broadcasting.* New York: Van Nostrand Reinhold.

National Association of Broadcasters. 1970. *Standard Definitions of Broadcast Research Terms.* Washington, D.C.: National Association of Broadcasters.

CHAPTER 15

Research in Advertising and Public Relations

For many years, research was not widely used in advertising and public relations; decisions were made on a more or less intuitive basis. However, with increased competition, mass markets, and mounting costs, more and more advertisers and public relations specialists have come to rely on research as a basic management tool.

Much of the research in advertising and public relations is **applied research,** which attempts to solve a specific problem and is not concerned with theorizing or generalizing to other situations. Advertising and public relations researchers are concerned with answering questions such as: "Should a certain product be packaged in blue or red?" "Is *Cosmopolitan* a better advertising buy than *Vogue*?" and "Should a company stress its minority hiring program in a planned publicity campaign?"

Advertising and public relations research does not involve any special techniques; the methods discussed earlier—laboratory, survey, field research, and content analysis—are in common use. They have been adapted, however, to provide specific types of information that meet the needs of these industries.

This chapter discusses the more common areas of advertising and public relations research and the types of studies they entail. In describing these research studies, the primary aim is to convey those facts that the reader must know in order to understand the methods involved and to use them intelligently. A significant portion of the research in these areas involves market studies conducted by commercial research firms; these studies form the basis for much of the more specific research that follows in either the academic or private sectors. While recognizing the importance of market research, this chapter does not have the space required to treat this topic. Readers who desire additional information about market research techniques should consult Tull and Hawkins (1976) and Ferber (1974).

Message Research

At the risk of oversimplification, there are two main areas of advertising research. **Message research,** also known as *copy testing,* involves studying advertisements or commercials in order to construct an advertisement or an entire campaign that will have a maximum impact on the target audience. Media research, which is treated in the following section, is used to discover the most efficient medium or combination of media for advertising a particular product or service.

Message research takes place in every stage of the advertising process. Before a campaign begins, it is used to determine which topics to stress and which to avoid. More research is conducted while the advertisement is being written, drawn, filmed, or photographed. Finally, it is employed to discover whether the campaign has achieved its prescribed goals. In short, message research helps advertisers to answer three basic questions: (1) "What should the message say?" (2) "How should it be said?" and (3) "What effect(s) did the message have?"

Message Content

Determining what the message should say is a necessary preliminary step in conducting an advertising campaign. For example, an advertiser is hired to promote a new brand of dog food. The manufacturer wishes to stress many different aspects of the product in the advertisement, such as cost, convenience, taste, and nutritional value. Due to time and space limitations, however, not all of these points can be included. The advertiser can use message research to determine which will be the most effective in selling the product. A content analysis of competitors' commercials, for example, might provide useful information. Also, a focus group interview of pet owners could help to determine what consumers look for when they purchase dog food; this could be followed by a more detailed survey of the pet owner population to ensure that the buying habits of the focus group are representative. Once the data from these studies have been collected and analyzed, the advertiser can make a more intelligent decision as to which points to include or stress in the campaign.

Roy and Nicolich (1980) conducted studies of this sort in order to isolate those attributes that customers felt were important in a retail store; they determined that customer satisfaction, honest advertising, and low-priced merchandise were the most important factors to stress in retail advertising. Myers (1970) conducted a regression analysis using consumer evaluations of a new snack food and found the important aspects to be taste, appearance, and strength of flavor. Corley (1970) did a similar analysis for an instant milk product, which suggested that the product's superiority to skim milk, its convenience, and its low price should be emphasized in advertising.

Message Structure

Once the content of an advertisement or commercial has been established, tests must be performed to ascertain the most effective way to structure these ideas. This type of research can take several forms. One widely used technique, known as **advertising element research,** tests each element of an advertisement separately to determine its effectiveness. For example, the illustration, headline, and body copy of a magazine spread might each be studied. The researcher might show each of two or more groups of subjects an illustration of a product photographed from a different angle to determine if the angle significantly affects product recall. He or she might evaluate the headline by allowing potential buyers to rate different typefaces used, and copy could be tested for comprehensibility, readability, and recall. The following studies illustrate this area of research.

Friedman, Termini, and Washington (1976) employed element research to test the relative effectiveness of four different endorsements for a new wine product. They found that each of the four endorsers produced a similar effect on consumers' perceptions of the product's selling price and of the advertisement's credibility; endorsements in and of themselves seemed to have a favorable impact on the subjects' intent to purchase the product. In a similar study, LaChance, Chestnut, and Lubitt (1977) found that the presence of a decorative female model in a printed advertisement aided in the recall of facts about the model but not of brand names.

A second type of research used to structure advertising content is known as a **complete execution,** so named because it allows an entire advertisement or commercial to be studied at one time. This procedure is often employed in broadcasting due to the difficulty of separating the components of a television or radio commercial. Particularly in the case of television, where the cost of producing a commercial can be exorbitant, many researchers prepare preliminary or rough versions for testing. This can be accomplished by means of a **storyboard,** a device that displays photographs or artists' renditions of the key scenes in a planned commercial as well as the accompanying dialog. Two or more storyboards can be constructed and shown to different experimental groups to gauge their effectiveness.

A more elaborate approach is to create a **rough commercial,** similar to the actual commercial except that amateur actors are used, locations are simplified, and the editing and narration often lack smoothness. In short, rough commercials are a "first draft" of the finished product. Several such commercials can be filmed in a single day using portable television equipment, thus providing a low-cost means of comparing different appeals and persuasive techniques. Some researchers argue that rough commercials are an ineffective means of evaluating structure. However, others argue that final commercials do not provide enough additional information to justify their expense. Also, few advertisers are willing to bear the expense of refilming a commercial even if it is shown to be ineffective.

Message Effectiveness

The final phase of message research is a posttest, initiated shortly after the first advertisement or commercial appears, to determine if the campaign is having the desired effects. It also allows any negative effects to be corrected before serious damage can be done to a company's sales, public image, and so forth.

Posttesting requires precisely defined campaign goals. Some campaigns, for example, are designed to draw customers away from competitors; others are conducted for the purpose of retaining a company's own customers. Still others are intended to enhance the image of a firm and may not be concerned with customers' purchase preferences. Each of these aims calls for a different measure of advertising effectiveness.

Two general types of research occur during the posttesting phase: (1) calculations of frequency of exposure and recall; and (2) studies of changes in awareness, perception, and purchase behavior. Suppose an advertiser has developed a campaign to promote a new brand of breakfast cereal. Using the first type of research, he or she might posttest the effectiveness of the campaign by surveying a sample of consumers and asking them if they remember seeing any commercials for breakfast cereals recently. If a large number recall seeing the commercial for the new cereal, the advertiser might conclude that the campaign has at least succeeded in making people aware of the product. This does not necessarily mean that they are buying the product, however; this determination must be made by using the second type of research. Thus, the advertiser might conduct a study to measure how consumers perceive the new cereal and how many have actually purchased it.

In general, advertisers are more concerned with the latter type of research, although exposure and recall figures can be helpful in planning future marketing efforts. In fact, **recall studies** are the most commonly used form of advertising research. In its simplest form, a recall study involves showing subjects a copy of a newspaper or magazine and asking them which advertisements they remember seeing or reading. The results are then used to tabulate a "reader traffic score" for each advertisement. This method is prone to criticism, however, because respondents might confuse the advertisements or the publications in which they were seen, or they might report having seen more than they actually did in order to please the interviewer. To control this problem, researchers often make use of aided recall techniques; for instance, they might also show the respondents a list of advertisers, some of whose advertisements actually appeared in the publication and some of whose did not.

For obvious reasons, this type of recall study is not entirely suitable for radio and television commercials; a more commonly used method in such cases is the telephone survey. Respondents are telephoned shortly after the commercial of interest is broadcast and asked if they were watching television or listening to the radio at that time; if the answer is yes, they are then asked what commercials they recall seeing or hearing.

In performing the second type of posttesting—evaluating consumer attitudes, perceptions, and purchasing behavior—researchers can choose from a wide variety of methods. One of the simplest and most common procedures is to expose one group of subjects to an advertising campaign and then compare it to a control group that has not been exposed. In this manner, a researcher can draw inferences as to the actual effects of the campaign on consumer attitudes.

Another important posttesting method is used to measure brand preference or buying predisposition. Typically, subjects are asked a question along these lines: "If you were going shopping tomorrow to buy breakfast cereal, which brand would you buy?" This might be followed by: "Would you consider buying any other brands?" and "Are there any cereals you would definitely not buy?" (This last question is included in order to determine whether the advertising campaign has had any negative effects.) Additionally, some researchers (Haskins 1976) suggest using a buying intention scale and instructing respondents to check the one position on the scale that best fits their intention. Such a scale might look like this:

_____ I'll definitely buy this cereal as soon as I can.
_____ I'll probably buy this cereal sometime.
_____ I might buy this cereal, but I don't know when.
_____ I'll probably never buy this cereal.
_____ I wouldn't eat this cereal even if somebody gave it to me.

The scale allows advertisers to see how consumers' buying preferences change during and after the campaign.

Perhaps the most reliable methods of posttesting are those that measure actual sales, direct response, and other easily quantifiable behavior patterns. Direct response, for example, might be measured by counting the number of redeemable coupons that are returned or the number of requests for free samples. Actual sales figures can be obtained in many ways. It is possible to measure sales indirectly, by asking a question such as: "What brand of breakfast cereal did you most recently purchase?" However, the findings would be subject to error due to faulty recall, courtesy bias, and so forth; for this reason, more direct methods are generally preferred. If enough time and money are available, direct observation of people's selections in the cereal aisles at a sample of supermarkets can be a useful source of data. Store audits that list the total number of boxes sold at predetermined times are another possibility. Last, and possibly most expensive, is the household audit technique, whereby an interviewer visits the homes of a sample of consumers and actually inspects their kitchen cupboards to see what brands of cereals are there. In addition to the audit, a traditional questionnaire is used to gather further information about the respondent's feelings toward the commercials.

Advertising research literature provides many good examples of posttest studies. Becknell and McIsaac (1963) evaluated the effects of a television advertising campaign; they found that the commercials increased the total cookware market by 21% and doubled the sales of the product advertised. Hoofnagle (1963) conducted a field experiment to test the impact of an advertising campaign for lamb. By monitoring purchases of lamb in a sample of supermarkets, Hoofnagle was able to show that a program of cooperative advertising raised sales by 26%. More recently, Bogart, Tolley and Orenstein (1970) conducted a posttest of 31 separate newspaper advertisements for several different products. They compared the purchasing behavior of subjects who had seen the advertisements with the behavior of a control group which had not. On balance, the advertisements were found to be effective.

Message Research by Private Firms

Because message analyses generally involve a great deal of time and resources, advertisers often hire professional research firms to conduct them. One such firm, the AdTel Company, uses special cable television facilities to test the effectiveness of commercials. The AdTel cable system, which is operated in three separate communities, is designed to carry two signals per channel, so that a household can receive either the "A" signal or the "B" signal. The households in these communities are grouped according to key demographic variables, and one group is connected to the A signal and the other to the B signal (the subscribers themselves do not know which channel they are receiving). The A households receive one version of the commercial being tested, while the B group receives either a different version or no commercial. About 1,000 families in each group keep diaries about the products and services that they purchase. The effectiveness of a commercial or campaign is judged by examining the purchasing behavior of the families in each group, as reflected in their diary entries. Measurement bias is eliminated because the families do not realize the connection between the diary and their television viewing. The AdTel system also has the advantage of measuring actual purchasing behavior as the dependent variable.

The Burke Marketing Research Company also offers a service, called the "Day-After-Recall Report," to measure the effectiveness of television commercials. Burke has the capacity to test recall of commercials by consumers in more than 30 cities, although typically only three to five cities are used in a single test. On the day after the commercial is run, the Burke interviewers conduct a telephone survey to obtain a sample of about 200 people, all of whom have watched the program that contained the commercial. These individuals are asked if they remember seeing the commercial and, if so, what details they can remember about it. Over the years, Burke has compiled data on numerous commercials and has computed average scores for recall; the scores (called "Burkes") of the commercials that the firm now tests can be considered in light of these norms, thus providing the advertiser with a benchmark for comparison.

Another method of posttesting television commercials is provided by Gallup and Robinson. Like Burke, Gallup and Robinson measure the percentage of respondents who remember seeing the commercial and of those who can remember specific points. Additionally, they provide a score indicating the degree of favorable attitude toward the product, based on positive statements made by the subjects during the interview.

Gallup and Robinson also conduct pretests and posttests of magazine advertisements. Their Magazine Impact Research Service (MIRS) measures recall of advertisements appearing in general interest magazines. Copies of a particular issue containing the advertisement under study are mailed to approximately 150 readers. (In the case of a pretest, the MIRS binds the proposed advertisement into each magazine.) The day after delivery of the magazines, respondents are telephoned and asked which advertisements they noticed in the magazine and what details they can remember about them. These results are then reported to the advertiser.

One of the most well-known professional research firms is the Starch Message Report Service, which conducts posttest recall and recognition research. Starch routinely measures advertising readership in more than 100 magazines and newspapers. Using a sample of approximately 300 people, the Starch interviewers take a copy of the periodical under study to the respondents' homes. If a subject has already looked through that particular magazine or newspaper, he or she is then questioned at length. The interviewer shows the respondent an advertisement and asks if he or she has seen or read any part of it. If the answer is no, the interviewer moves on to another advertisement; if the answer is yes, more questions are asked to determine how much of it was read. This procedure continues until the respondent is questioned about every advertisement in that issue up to 100 (at which point the interview is terminated to avoid subject fatigue). Starch then places each respondent into one of four categories for each advertisement:

1. *Non-Reader* (a person who did not recall seeing the advertisement)
2. *Noted Reader* (a person who remembered seeing the advertisement)
3. *Associated Reader* (a person who not only saw the advertisement but also read some part of it that clearly indicated the brand name)
4. *Read Most Reader* (a person who read more than half of the written material in the advertisement)

The Starch organization reports the findings of its recall studies in a novel manner. Advertisers are given a copy of the magazine in which readership scores printed on yellow stickers have been attached to each advertisement. Figure 15.1 is an example of a "Starched" advertisement.

The Starch Message Report provides a measurement of recognition only; for an indication of an advertisement's success in getting its message across, advertisers can request that a Starch Reader Impression Study be performed. These studies also involve in-depth interviews with readers; those who have seen an advertisement in a particular newspaper or magazine are asked a series of detailed questions about it, such as:

1. "In your own words, what did this ad tell you about the product?"
2. "What did the pictures tell you?"
3. "What did the written material tell you?"

The responses then undergo content analysis, and the results are summarized for clients. Additionally, Starch reports the percentage of favorable and unfavorable comments about each advertisement.

One final company that conducts message research is the Advertising Index, which provides posttest data on brand awareness and advertising effectiveness. The Index collects these data through telephone interviews with respondents selected randomly from particular geographic areas. The interviews are lengthy and include questions designed to test such variables as usage of certain products, brands of products recently purchased, and effectiveness of specific advertisements. For example, one measure of effectiveness, known as "Top-of-Mind Awareness," is based on responses to the question: "When you think of (*a product*), what brand first comes to mind?" Questions regarding advertisement recall are also asked in order to compare the purchasing behavior of subjects who remember seeing a particular advertisement with that of those who do not. The Index performs such studies for both print and electronic media advertising campaigns.

Figure 15.1 Starched ad

Source Starch INRA Hooper, Inc Readership Report of *Southern Living*, April 1977, p 77 Reprinted by permission

Media Research

The basic goal of media research is to determine the most efficient advertising vehicle for a particular product or service. Media research studies are designed to measure advertising **media efficiency,** the most efficient medium or combination of media being that which reaches the largest number of potential consumers at the lowest cost. As mentioned in Chapter 14, a common measure of media efficiency is the cost-per-thousand (CPM) ratio. In advertising research, the CPM is computed by the following formula:

$$\text{CPM} = \frac{\$\text{ Cost of Ad}}{\text{audience size} \div 1{,}000}$$

Thus, if an advertisement that costs $40,000 reaches an audience of 400,000 persons, its CPM would be $100.

Advertisers generally conduct media research for one of three reasons: (1) to determine the audience size and composition of a medium or media; (2) to study the relative frequencies of advertising exposure provided by various combinations of media; and (3) to obtain information about the advertising activity of competitors. The studies that are conducted in each of these areas can provide valuable information with regard to media efficiency.

Audience Size and Composition

Analyses of audiences are probably the most commonly used advertising studies in both print and electronic media research. Since advertisers spend large amounts of money in the print and electronic media, it is understandable that advertisers are very interested in the audiences for those messages. In most cases, audience information is gathered using techniques that are compromises between what is practical and what is ideal.

The audience size of a newspaper or magazine is commonly measured in terms of the number of copies distributed per issue. This number, which is called the publication's **circulation,** includes all copies delivered to subscribers as well as those bought at newsstands or from other sellers. Because a publication's advertising rate is directly determined by its circulation, the print media have developed a standardized method of measuring circulation and have instituted an organization, the Audit Bureau of Circulation (ABC), to verify that a publication actually distributes the number of copies per issue that it claims. (The specific procedures used by the ABC are discussed later in this chapter.)

Circulation figures are used to compute the CPMs of various publications. For example, suppose Newspaper X charges $1,800 for an advertisement and has an ABC-verified circulation of 180,000;

Newspaper Y, with a circulation of 300,000, charges $2,700 for the same size space. Which newspaper is more efficient?

	Newspaper X	Newspaper Y
Ad cost	$1,800	$2,700
Circulation	180,000	300,000
Cost per thousand circulated copies	$\frac{\$1{,}800}{180} = \$10.00$	$\frac{\$2{,}700}{300} = \$9.00$

Thus, Newspaper Y is the more efficient advertising vehicle.

Note that this method only considers the number of circulated copies of a newspaper or magazine. While this information is useful, it does not necessarily indicate the total number of readers of the publication. To estimate the total audience, the circulation figure must be multiplied by the average number of readers of each copy of an issue. This information is obtained by performing audience surveys.

A preliminary step in conducting such surveys is to operationally define the concept *magazine reader* or *newspaper reader.* There are many possible definitions, but the one most commonly used is fairly liberal: a *reader* is a person who has read or at least looked through an issue.

Three techniques are used to measure readership. The most rigorous is the unaided recall method, in which respondents are asked if they have read any newspapers or magazines in the past month (or other time period). If the answer is yes, subjects are asked to specify the magazines or newspapers they read. When a publication is named, the interviewer attempts to verify reading by asking questions about the contents of that publication. The reliability of the unaided recall method is open to question (as has been discussed) because of the difficulty respondents often have in recalling specific content.

A second technique involves aided recall. In this method, the interviewer names several publications and asks if the respondent has read any of them lately. Each time the respondent claims to have read a publication, the interviewer asks if he or she remembers seeing the most recent copy. The interviewer may jog a respondent's memory by describing the front page or the cover. Finally, the respondent is asked to recall anything that was seen or read in that particular issue. (A variation on this process is called **masked recall:** respondents are shown the front page or the cover of a publication with the name blacked out and are asked if they remember reading that particular issue. If the response is affirmative, they are then asked to recall any items they have seen or read.)

The third technique is called the **recognition method.** It entails showing respondents the logo or cover of a publication. For each publication that the respondent has seen or read, the interviewer produces a copy and the respondent leafs through it to see which articles or stories he or she recognizes. All respondents who definitely remember reading the publication are counted in its audience. To check the accuracy of the respondent's memory, dummy articles may be inserted into the interviewer's copy of the publication; respondents who claim to have read the dummy items may thus be eliminated from the sample or given less weight in the analysis. Many advertising researchers consider the recognition technique to be the most accurate predictor of readership scores.

Once the total audience for each magazine or newspaper has been tabulated, the advertiser can determine which publication is the most efficient buy. For example, returning to the above example, suppose Newspaper X and Newspaper Y have the following audience figures:

	Newspaper X	Newspaper Y
Ad cost	$1,800	$2.700
Circulation	180,000	300,000
CPM	$10.00	$9.00
Number of people who read the issue	630,000 (3.5 readers per copy)	540,000 (1.8 readers per copy)
Revised CPM	$2.86	$5.00

On the basis of these figures, Newspaper X is seen to be the more efficient choice.

Another variable to be considered in determining the advertising efficiency of a newspaper or magazine is the number of times a person reads each issue. For example, imagine a situation in which two newspapers or magazines have exactly the same number of readers per issue. Publication A consists primarily of pictures and contains little text; people tend to read it once and then not look at it again. Publication B, on the other hand, contains several lengthy and interesting articles; people pick it up several times. It seems reasonable that Publication B would be a more efficient advertising vehicle, since it provides several possible exposures to an advertisement for the same cost as Publication A. Unfortunately, a practical and reliable method for measuring the number of exposures per issue has yet to be developed.

Perhaps the most important gauge for advertising efficiency is the composition of the audience. It matters little if an advertisement for farm equipment is seen by 100,000 people if only a few of them are in the market for such products. In order to evaluate the number of potential customers in the audience, an advertiser must first conduct a survey to determine certain demographic characteristics of the people who tend to purchase a particular product. For example, potential customers for beer might be typically described as males between the ages of 18 and 49; those for fast-food restaurants might be households in which the primary wage earner is between 18 and 35 and there are at least two children under 12. These demographic characteristics of the "typical consumer" are then compared with the characteristics of a publication's audience (also determined by a survey) in order to tabulate the potential buying audience for the product. The cost of reaching this audience is also expressed in CPM units:

	Newspaper X	Newspaper Y
Ad cost	$1,800	$2,700
Circulation	180,000	300,000
CPM	$10.00	$9.00
Number of people who read average issue	630,000	540,000
Number of potential beer drinkers	150,000	220,000
Number of potential fast food customers	300,000	200,000
CPM (beer drinkers)	$12.00	$12.27
CPM (fast food customers)	$ 6.00	$13.50

An examination of these figures indicates that Newspaper X is slightly more efficient as a vehicle for reaching potential beer customers and much more efficient in reaching fast-food restaurant patrons.

Determining audience size and composition in the electronic media poses special problems for advertising researchers, due to the ephemeral nature of radio and television broadcasts. For a detailed discussion of the techniques involved in this type of audience or ratings research, the reader is referred to Chapter 14.

Frequency of Exposure in Media Schedules

An advertiser who is working within a strict budget to promote a product or service may be limited to the use of a single vehicle or medium. Often, however, an advertising campaign is conducted via several advertising vehicles simultaneously. The question then arises as to which combination of vehicles and/or media provides the greatest frequency of exposure for the advertiser's product.

A substantial amount of recent media research has been devoted to this question. Much of this research has concentrated on the development of mathematical models of advertising media and their audiences. The actual mathematical derivations of these models are beyond the scope of this book. The following material provides simplified descriptions of the concepts underlying the models; readers who wish to pursue the subject in more detail should consult Aaker and Myers (1975).

Central to an understanding of these media models is the notion of **objective function,** which is a mathematical statement that provides a quantitative value for a given media combination (also known as a *schedule*). This value represents the schedule's effectiveness in providing advertising exposure. The advertising researcher's task is to calculate the objective functions of various media schedules in order to determine which schedule offers the maximum exposure for his or her product.

Calculations of objective function are based on values generated by studies of audience size and composition for each vehicle or medium. In addition, a schedule's objective function value takes into account such variables as the probability that the advertisement will be forgotten; the total cost of the media schedule, as it compares with the advertiser's budget; and the "media option source effect" (that is, the relative impact of exposure in a particular advertising vehicle—for example, an advertisement for men's clothes is likely to have more impact in *Gentlemen's Quarterly* than in *True Detective*).

Two computer media models are frequently used to calculate objective functions. The first is MEDIAC (Little and Lodish 1969), which is designed to maximize sales in a particular market segment by allowing a product's market potential within that segment to be included as an additional variable. The MEDIAC program calculates the probability that a given person within a market segment will be exposed to an advertisement. This probability depends on such factors as the size of the advertisement, the use of color, and the characteristics of the media vehicle used to deliver the message. It is assumed that exposure to one or more advertisements will affect a person's willingness to buy the advertised product. The advertiser provides MEDIAC with data on the probable sales response to different levels of exposure within all relevant segments of the market; MEDIAC computes the probable sales response for each market segment and then totals the response for the entire audience for any given schedule.

A second computer model, called ADMOD (Aaker 1975), is designed to maximize favorable attitude changes among consumers toward the advertised product. ADMOD evaluates a media schedule by examining its likely impact on each individual in samples drawn from the market population. This impact is calculated by taking into account the number and source of exposures for each individual and the effect of these exposures on the probability of obtaining the desired attitude change. The results are then projected to the population. A unique ADMOD feature allows the researcher to include certain data about different message strategies that might be employed. As media schedules increase in complexity and become more expensive, it is expected that the use of computer models such as MEDIAC and ADMOD will increase.

Competitors' Activities

It is often helpful to advertisers to know the advertising media choices of their competitors. This information can help the advertiser to avoid making the mistakes of less successful competitors and to imitate the strategies of the more successful. Moreover, if an advertiser seeking to promote a new product knows that the three leading competitors are using basically the same media mix, he or she might feel that their consensus is worthy of consideration.

It is possible for an advertiser to collect data on competitors' activity either by setting up a special research team or by subscribing to the services of a syndicated research company. Since the job of monitoring the media activity of a large number of firms advertising in several media is so difficult, most advertisers rely on the syndicated service. Such services gather data by direct observation, that is, by tabulating the advertisements that appear in a given medium. For some media, such as magazines, television, and radio, the complete population of advertisements is studied. For others, such as newspapers, samples are observed and the results generalized. In addition to information about frequency of advertisements, cost figures are also helpful; these estimates are obtained from the published rate cards of the various media vehicles.

Advertisers also find it helpful to know *what* competitors are saying. To acquire this information, many advertising agencies conduct systematic content analyses of the messages in a sample of the competitors' advertisements. The results often provide insight into the persuasive themes, strategies, and goals of competitors' advertising. Such studies are the reason many commercials tend to look and sound alike: successful commercial approaches are often mimicked.

Media Research by Private Firms

As mentioned earlier, the Audit Bureau of Circulation (ABC) supplies advertisers with data on the circulation figures of newspapers and magazines. As of 1980, ABC measures the circulation of about 75% of all print media vehicles in the U.S. and Canada. ABC requires publishers to submit a detailed report of their circulation every six months; it verifies these reports by having its field workers visit each publication to conduct an audit. The auditors typically examine records of the publications' press runs, newsprint bills, or other invoices for paper, as well as transcripts of circulation records and other related files.

The ABC audit results, as well as overall circulation data, coverage maps, press times, and market data, are published in an annual report and distributed to ABC members and advertisers. ABC now reports data on audience size for certain selected newspapers. Called the "Newspaper Audience Research Data Bank," this report consists of a collection of audience surveys conducted by newspapers in the top 100 markets.

Verified Audit Circulation (VAC) is another organization that audits the circulation of newspapers and magazines. While ABC measures only paid circulation, VAC tabulates both free and paid circulation figures. (Free circulation publications are those that are given away; for example, shoppers' bulletins and magazines given to airline passengers.) It uses procedures somewhat similar to those of ABC in its newspaper audits, while magazine audits are conducted by counting the names on a publication's subscriber list and by mailing questionnaires to a sample of subscribers.

The Simmons Market Research Bureau provides comprehensive feedback about magazine readership. This service selects a large random sample of readers and shows them illustrations of the titles of about 70 magazines to determine which ones they have recently read or looked through. Subjects are then shown stripped-down versions of the publications they identified, and readership is verified by further questioning. At the same time, data are gathered about the ownership, purchase, and use of a wide variety of products and services. This information is tabulated by Simmons and released in a series of detailed reports which contain specific information regarding the demographic makeup and purchasing behavior of each magazine's audience. Using these data, advertisers can determine the cost of reaching potential buyers of their products or services. A portion of a Simmons Report is reproduced in Table 15.1.

Two companies—Arbitron and A. C. Nielsen—supply broadcast audience data for advertisers. Arbitron measures radio listening in about 160 markets across the United States, while both Arbitron and Nielsen provide audience estimates for local television markets. Both companies use the diary technique to collect data, and both use essentially the same procedures for selecting homes into their sample. (For more information on the methods employed by these two companies and others, see the discussion in Chapter 14.)

Private research firms are also available to provide advertisers with information about their competitors' advertising campaigns.

Data about magazine advertising are collected by *Leading National Advertisers* (LNA), in cooperation with the Publishers Information Bureau (PIB). The publishers of magazines that belong to the PIB mark all paid advertising in each issue and send the marked copies to LNA, where trained coders record detailed information about each advertisement. This information is then recorded in a report sent to LNA subscribers. The data are arranged according to product type and brand name. By scanning the LNA reports, it is possible to determine which magazines competitors are using, the size of the advertisements they purchase, when they appear, and their approximate cost.

Information about advertising activity in newspapers is supplied by Media Records. Since it is impossible to examine every newspaper in the country, Media Records concentrates on about 225 newspapers published in approximately 75 urban areas. During a specified time period, coders measure the size of every advertisement appearing in each issue of these papers and also estimate their cost. The data from this sample are then used to project estimates for the top 125 market areas. Media Records publishes this information in its

Table 15.1. Example of Simmons report

			Broadminded				Creative				Dominating				Efficient			
		Total U.S. '000	A '000	B % Down	C Across %	D Index	A '000	B % Down	C Across %	D Index	A '000	B % Down	C Across %	D Index	A '000	B % Down	C Across %	D Index
Total Adults		158437	61916	100.0	39.1	100	43842	100.0	27.7	100	61408	100.0	38.8	100	49042	100.0	31.0	100
Magazines	Quintile 1	31799	14393	23.2	45.3	116	10510	24.0	33.1	119	14280	23.3	44.9	116	10764	21.9	33.9	109
	Quintile 2	32954	13592	22.0	41.2	106	9897	22.6	30.0	109	13319	21.7	40.4	104	10637	21.7	32.3	104
	Quintile 3	26611	10309	16.6	38.7	99	7394	16.9	27.8	100	9856	16.1	37.0	96	8672	17.7	32.6	105
	Quintile 4	31322	11819	19.1	37.7	97	8254	18.8	26.4	95	11811	19.2	37.7	97	9632	19.6	30.8	99
	Quintile 5	35752	11802	19.1	33.0	84	7787	17.8	21.8	79	12143	19.8	34.0	88	9337	19.0	26.1	84
Newspapers Weekday	Tercile 1	25455	10839	17.5	42.6	109	7254	16.5	28.5	103	10383	16.9	40.8	105	8790	17.9	34.5	112
	Tercile 2	67522	26639	43.0	39.5	101	18433	42.0	27.3	99	25971	42.3	38.5	99	21484	43.8	31.8	103
	Tercile 3	65460	24438	39.5	37.3	96	18154	41.4	27.7	100	25054	40.8	38.3	99	18768	38.3	28.7	93
TV-Prime Time	Quintile 1	31001	12033	19.4	38.8	99	8352	19.1	26.9	97	11273	18.4	36.4	94	9665	19.7	31.2	101
	Quintile 2	31417	12330	19.9	39.2	100	8180	18.7	26.0	94	12493	20.3	39.8	103	9407	19.2	29.9	97
	Quintile 3	32124	12750	20.6	39.7	102	9311	21.2	29.0	105	13299	21.7	41.4	107	10445	21.3	32.5	105
	Quintile 4	32131	12493	20.2	38.9	99	9088	20.7	28.3	102	12431	20.2	38.7	100	10018	20.4	31.2	101
	Quintile 5	31764	12309	19.9	38.8	99	8912	20.3	28.1	101	11912	19.4	37.5	97	9507	19.4	29.9	97
TV-Daytime	Tercile 1	45111	17747	28.7	39.3	101	11262	25.7	25.0	90	16297	26.5	36.1	93	13056	26.6	28.9	94
	Tercile 2	45241	17258	27.9	38.1	98	12093	27.6	26.7	97	17441	28.4	38.6	99	13426	27.4	29.7	96
	Tercile 3	68084	26911	43.5	39.5	101	20488	46.7	30.1	109	27671	45.1	40.6	105	22560	46.0	33.1	107
TV Total	Quintile 1	31204	12203	19.7	39.1	100	7668	17.5	24.6	89	11474	18.7	36.8	95	8928	18.2	28.6	92
	Quintile 2	31981	11953	19.3	37.4	96	8306	18.9	26.0	94	12352	20.1	38.6	100	9512	19.4	29.7	96
	Quintile 3	31816	13041	21.1	41.0	105	9066	20.7	28.5	103	11970	19.5	37.6	97	10421	21.2	32.8	106
	Quintile 4	31404	12396	20.0	39.5	101	9662	22.0	30.8	111	12702	20.7	40.4	104	10062	20.5	32.0	104
	Quintile 5	32033	13223	19.9	38.5	98	9140	20.8	28.5	103	12911	21.0	40.3	104	10121	20.6	31.6	102

Source: Simmons Market Research Bureau, *1980 Study of Media & Markets*, Volume M-5, pp 0368, 0369 Reprinted by permission

"Bluebooks," which provide data on advertising activity for all comparable products and services.

The most comprehensive information about advertisers' activities and expenditures is provided by Broadcast Advertisers Reports (BAR). This organization collects data on commercials appearing on radio and television networks as well as on local television stations. Advertising activity on network radio is measured by recording all network programs on audio tape. The tapes are played back for coders, who record programs, lengths of commercials, advertising companies, brand names, and other details. BAR estimates the cost of each commercial by referring to the network's published rate card. Totals are computed for each brand, and the results are published in BAR's quarterly "Network Radio Reports." The reports allow an advertiser to determine competitors' schedules and expenditures.

BAR uses basically the same technique to compile its weekly "Network Television Reports." Network programs are recorded and all commercials coded according to brand, length, and program. Rate cards are used to estimate costs. The report also contains a daily log of advertising, indexed by product type, and a total of the week's advertising costs per product.

BAR monitors advertising on local television stations by sampling approximately 275 stations located in the top 75 television markets. In the two top markets, New York and Los Angeles, measurements are made every week. In the other 73 markets, BAR selects one week per month for analysis, during which all station programming is recorded or videotaped, and this sample is used to make projections for the entire month. BAR issues two summary reports, a local market report for each city and a cumulative report covering all 75 markets. The types of data and the format are similar to those used for the network summaries.

Local radio advertising is not monitored by BAR. The only data available for advertisers in this area are provided by Radio Expenditure Reports (RER), which sends questionnaires to about 800 stations in the top 150 markets. Each station is asked to provide an estimate of its advertising revenue generated from regional and national commercials. RER tabulates the data and projects a national figure for all such radio advertising. An RER report tells an advertiser how much a competitor spent for radio spots during a certain time period, but it does not show when or on what stations the competitor advertised.

Public Relations Research

Public relations, like advertising, has become more research-oriented in recent years. Techniques such as survey research, content analysis, and focus groups are now widely employed in this field. Public relations researchers, however, use these methods for a highly specific reason: to improve communication with various publics.

Uses of Public Relations Research

The most common use of research in public relations is to gather data on audience attitudes and opinions. Such surveys are often conducted before, during, and after a public relations campaign. They may be custom-designed by the researcher or provided by professional research organizations. The Opinion Research Corporation, Roper Public Opinion Research Center, and Louis Harris and Associates are some of the more well-known organizations that supply data to public relations practitioners.

Public relations research is also used as a means of surveillance. Specific types of survey and content research, which will be discussed in more detail below, serve as early warning systems to identify public relations problems before they occur. Many companies and organizations have public relations departments that systematically scan the environment in order to keep abreast of changes in society that might cause short-term or long-term problems.

Another common use of research by public relations departments is to secure management support for their own functions, policies, recommendations, and so forth. In many organizations, the marketing, financial, and production departments are the ones that can most effectively influence the decision-making process. Part of this effectiveness is due to the fact that these departments typically rely on factual data and evidence to support their point of view. If a public relations department wishes to argue for a different course of action, it will naturally take advantage of the research methods at its disposal in order to provide a factual basis for this argument.

Finally, public relations research is often conducted to evaluate the effectiveness of a planned communication program. The pretest/posttest arrangement, discussed in the section on message research, is equally appropriate to gauging the results of a public relations campaign.

Types of Public Relations Research

Lerbinger (1977) identified four major categories of public relations research: (1) environmental monitoring programs; (2) public relations audits; (3) communication audits; and (4) social audits.

Researchers use **environmental monitoring programs** to observe trends in public opinion and social events that may have a significant impact on an organization. Generally, two different phases are involved. The first is the "early warning" phase: an attempt to identify emerging issues. This often takes the form of a systematic content analysis of publications likely to herald new developments. For example, one corporation has conducted a content analysis of scholarly journals in the fields of economics, politics and science; another company sponsors a continuing analysis of trade and general newspapers.

An alternate method is to perform panel studies of community leaders or other influential and knowledgeable citizens. These individuals are surveyed regularly with regard to the ideas they perceive to be important, and the interviews are analyzed to pick out new topics of interest.

The second phase of environmental monitoring consists of tracking public opinion on major issues. Typically this involves either a longitudinal panel study, in which the same respondents are interviewed several times during a specified interval, or a cross-sectional public opinion poll, in which a random sample is surveyed only once. One good example of the latter type of monitoring is seen in the efforts of the Television Information Office. Since 1959, this organization has commissioned the Roper Company to survey public attitudes concerning the credibility of television and other media. These data give a clear picture of how public opinion has fluctuated over the years.

The **public relations audit,** as its name suggests, is a comprehensive study of the public relations position of an organization. Such studies are used to measure a company's standing both internally—in the eyes of its employees—and externally—with regard to the opinions of customers, stockholders, community leaders, and so on. In short, as summarized by Simon (1980), the public relations audit is a "research tool used specifically to describe, measure and assess an organization's public relations activities and to provide guidelines for future public relations programming."

The first step in a public relations audit is to list those segments of the public that are most important to the organization. This is generally accomplished through personal interviews with key management personnel in each department and by a content analysis of the company's external communications. The second step is to determine how the organization is viewed by each of these audiences. This involves conducting a corporate image study, that is, a survey of audience samples. The questions are designed to measure familiarity with the organization—can the respondents recognize the company logo, identify a product it manufactures, or remember the president's name?—as well as attitudes and perceptions toward it. Ratings scales are often used; for example, respondents might be asked to rank their perceptions of the ideal electric company on a seven-point scale between adjective pairs such as:

The Ideal Electric Company

good ____:____:____:____:____:____:____ bad

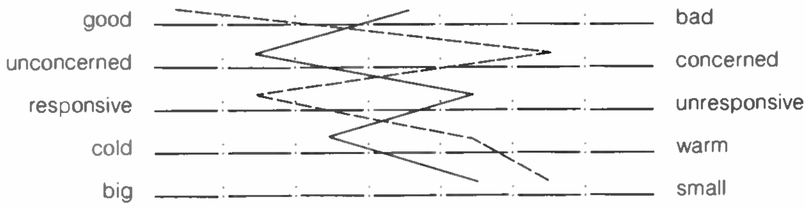
unconcerned ____:____:____:____:____:____:____ concerned

responsive ____:____:____:____:____:____:____ unresponsive

cold ____:____:____:____:____:____:____ warm

big ____:____:____:____:____:____:____ small

Later, the respondents could be asked to rate the specific electric company on the same scales. The average score for each scale is then tabulated, and the means are connected by a zigzag line to form a composite profile. Thus, in the graph below, the ideal electric company's profile is represented by a broken line and the actual electric company's standing by a solid line:



By comparing the two lines, public relations researchers can readily identify the areas in which a company falls short of the ideal. Corporate image studies can also be conducted before the beginning of a public relations campaign and again at the conclusion of the campaign to evaluate its effectiveness.

The **communication audit** resembles a public relations audit but has narrower goals; it concerns the internal and external means of communication used by an organization, rather than the company's entire public relations program. The three research techniques generally used in conducting such an audit are readership surveys, content analyses, and readability studies. Readership studies are designed to measure how many people read and/or remember certain publications (such as employee newsletters or annual reports) and the messages they contain. The results are used to improve the content, appearance, and method of distribution of the publications. Content analyses are used to see how the media are handling news and other information about and from the organization; they may be conducted in-house or by private firms that provide computerized studies of press coverage. Readability studies help a company gauge the ease with which its employee publications and press releases can be read.

A **social audit** is a small-scale environmental monitoring program designed to measure an organization's social performance, that is, how well it is living up to its social responsibilities. The audit provides feedback on such company-sponsored social action programs as minority hiring, environmental cleanup, and employee safety. This is the newest form of public relations research and also the most challenging. Researchers are currently studying such questions as what activities to audit, how to collect data, and how to measure the effects of the programs. Nevertheless, several large companies, including General Motors and Celanese, have already conducted lengthy social audits.

Summary

There are two main areas of advertising research: message research and media research. Message research consists of studies that examine an advertisement or commercial itself. It is used to determine what aspects of a product should be stressed in the advertising, to evaluate the effectiveness of the various components of the advertisement, and to discover if the campaign had the intended effects. Media research helps determine which advertising vehicles are the most efficient and what type of media schedule will have the greatest impact. Various techniques are available for measuring the advertising effectiveness of different media. Many private firms specialize in supplying message and media data to advertisers.

Research in public relations involves monitoring relevant developments and trends, studying the public relations position of an organization, examining the messages produced by an organization, and measuring how well an organization is living up to its social responsibilities.

Questions and Problems for Further Investigation

1. Assume you have just developed a new diet soft drink and are ready to market it. Develop a possible research study that you could use to identify the elements and topics you should stress in your advertising.
2. A full-page advertisement costs $16,000 in Magazine A and $26,000 in Magazine B. Magazine A has 2.5 readers per copy, while Magazine B has 1.8. In terms of CPM readers, which magazine is the most efficient advertising vehicle?
3. Select a sample of newspaper and magazine advertisements for two different airlines. Conduct a content analysis of the themes or major selling points in each advertisement. What similarities and differences are there?
4. Assume you are the public relations director for a major automobile manufacturer. How would you go about conducting an environmental monitoring study?

References and Suggested Readings

Aaker, D. 1975. "ADMOD, an Advertising Decision Model." *Journal of Marketing Research* 12:February, 37–45.

Aaker, D., and Myers, J. 1975. *Advertising Management.* Englewood Cliffs, N.J.: Prentice-Hall.

Anderson, R., and Barry, T. 1979. *Advertising Management: Text and Cases.* Columbus: Charles E. Merrill.

Aronoff, C. 1975. "Credibility of Public Relations for Journalists." *Public Relations Review* 1:2, 43–54.

Becknell, J., and McIsaac, R. 1963. "Test Marketing Cookware Coated with Teflon." *Journal of Advertising Research* 3:3, 2–8.

Bogart, L., Tolley, S., and Orenstein, F. 1970. "What One Little Ad Can Do." *Journal of Advertising Research* 10:4, 3–15.

Bowes, J., and Stamm, K. 1975. "Evaluating Communication with Public Agencies." *Public Relations Review* 1:1, 23–27.

Corley, L. 1970. "How to Isolate Product Attributes." *Journal of Advertising Research* 10:4, 41–46.

Cutlip, S., and Center, A. 1978. *Effective Public Relations*. Englewood Cliffs, N.J.: Prentice-Hall.

Dirksen, C., Kroeger, A., and Nicosia, F. 1977. *Advertising Principles, Problems and Cases*. Homewood, Ill.: Richard D. Irwin.

Ferber, R., ed. 1974. *Handbook of Marketing Research*. New York: McGraw-Hill.

Fletcher, A., and Bowers, T. 1979. *Fundamentals of Advertising Research*. Columbus: Grid Publishing Company.

Friedman, H., Termini, S., and Washington, R. 1976. "The Effectiveness of Advertising Using Four Types of Endorsers." *Journal of Advertising* 5:3, 22–24.

Haskins, J. 1976. *An Introduction to Advertising Research*. Knoxville: Communication Research Center.

Hoofnagle, W. 1963. "The Effectiveness of Advertising for Farm Products." *Journal of Advertising Research* 3:4, 2–6.

LaChance, C., Chestnut, R., and Lubitt, A. 1977. "The Decorative Female Model." *Journal of Advertising* 6:4, 11–14.

Lerbinger, O. 1977. "Corporate Use of Research in Public Relations." *Public Relations Review* 3:4, 11–20.

Little, J., and Lodish, L. 1969. "A Media Planning Calculus." *Operations Research* January-February, 1–35.

Myers, J. 1970. "Finding Determinants of Buying Attitudes." *Journal of Advertising Research* 10:6, 9–13.

Roy, R., and Nicolich, M. 1980. "PLANNER: A Media Market Position Model." *Journal of Advertising Research* 20:2, 61–68.

Simon, R. 1980. *Public Relations: Concepts and Practices*. Columbus: Grid Publishing Company.

Tull, D., and Hawkins, D. 1976. *Marketing Research*. New York: Macmillan.

PART 5

Analyzing and Reporting Data

CHAPTER 16

The Computer as a Research Tool

Introduction to Computer Analysis □ Data Input Equipment □ Using the Computer □ A Typical Computer Program □ Further Suggestions for Operators □ Summary □ Questions and Problems for Further Investigation □ References and Suggested Readings □

The use of computers in mass media research is no longer considered to be out of the ordinary. Only a few years ago, researchers were being told that high-speed computers would make it possible to analyze complex research problems quickly. Today, this prediction is a reality, and ongoing technological developments continue to make computers one of the most fascinating areas of research. The miniaturization of internal computer components has allowed manufacturers to replace the room-size computers of ten years ago with hand-held calculators and other incredibly small devices. Most media researchers now have some type of computer at their disposal, and the reliance upon computers as a research tool is more evident than at any other time in history.

There are many types of computers in existence, from the highly sophisticated models used by NASA and other scientific agencies to home computers available through various commercial outlets. The majority of mass media researchers use a general purpose computer located in a college, university, or business office. These computers are able to accept and execute a wide variety of programs, either sequentially or, in some cases, simultaneously (Klecker, Nie, and Hull 1975). The discussions in this chapter relate to the general purpose computer, although the basic principles relate to most other types of computers as well. The purpose of the chapter is to provide an introduction to computer use and terminology in order to eliminate the hesitancy with which some researchers view the computer.

Introduction to Computer Analysis

Most mass media researchers use pre-prepared computer packages that contain a wide variety of ready-to-use statistical programs or sets of instructions to the computer. These **canned computer programs** require only a series of control cards or statements to be operable. Some of the most popular canned programs include SPSS (Statistical Package for the Social Sciences—Nie, Bent, and Hull 1975), BMD and BMD-P (Biomedical Computer Programs—Dixon 1973), SAS (Statistical Analysis System—Barr and others 1976), and OSIRIS III (Organized Set of Integrated Routines for Investigation with Statistics—Inter-University Consortium for Political Research 1973).

Canned packages are stored on computer cards or on magnetic tapes, or disks. Users do not come in direct contact with the packages but rather "call them up" by means of instructions on computer cards or statements on a video terminal. The control cards or statements used to call up a specific package differ from one computer center to another, but the purpose is the same: to retrieve the package and "run" a specific program.

In some cases canned packages are not appropriate for a researcher's problem, and the creation of a unique program is required. This is normally done by a computer programmer and is both time-consuming and expensive. Fortunately, a sufficient variety of mass media research packages are available so that instances where unique programs must be developed are rare.

There are three basic steps involved in any computer analysis: (1) selection of a suitable package; (2) data input; and (3) data output. The first two steps are performed by the researcher; once they are accomplished, the computer's internal control and coordination are handled by its **central processing unit** (CPU), which decides the order in which programs are to be run. Normally, the computer organizes programs according to their respective levels of difficulty or the amount of time required to compute each. For many programs, the time involved will be well under 5 seconds. Even a complicated multivariate problem (see Chapter 12) involving a large sample and several variables may take only 10 or 15 seconds of CPU time. The computer example presented later in this chapter can be run in slightly over a second.

CPU time is important to a computer center, since charges for computer use are partially based on this time. It is also important for the researcher: programs that require only a few seconds of CPU time generally have a fast "turnaround," meaning that only a short time period elapses between input and output. The rate of turnaround depends on how many other jobs are in the computer as well as the complexity of the individual program.

Another part of the computer with which researchers should be familiar is the **central memory** (also known as *core storage*). The central memory is used to temporarily store the information necessary to solve a specific problem. Once the program is complete, the memory is reused for another program. Many computer centers require users to specify the amount of storage space required to compute a given program. This is a matter that must be discussed with the operators at a particular center.

Data Input Equipment

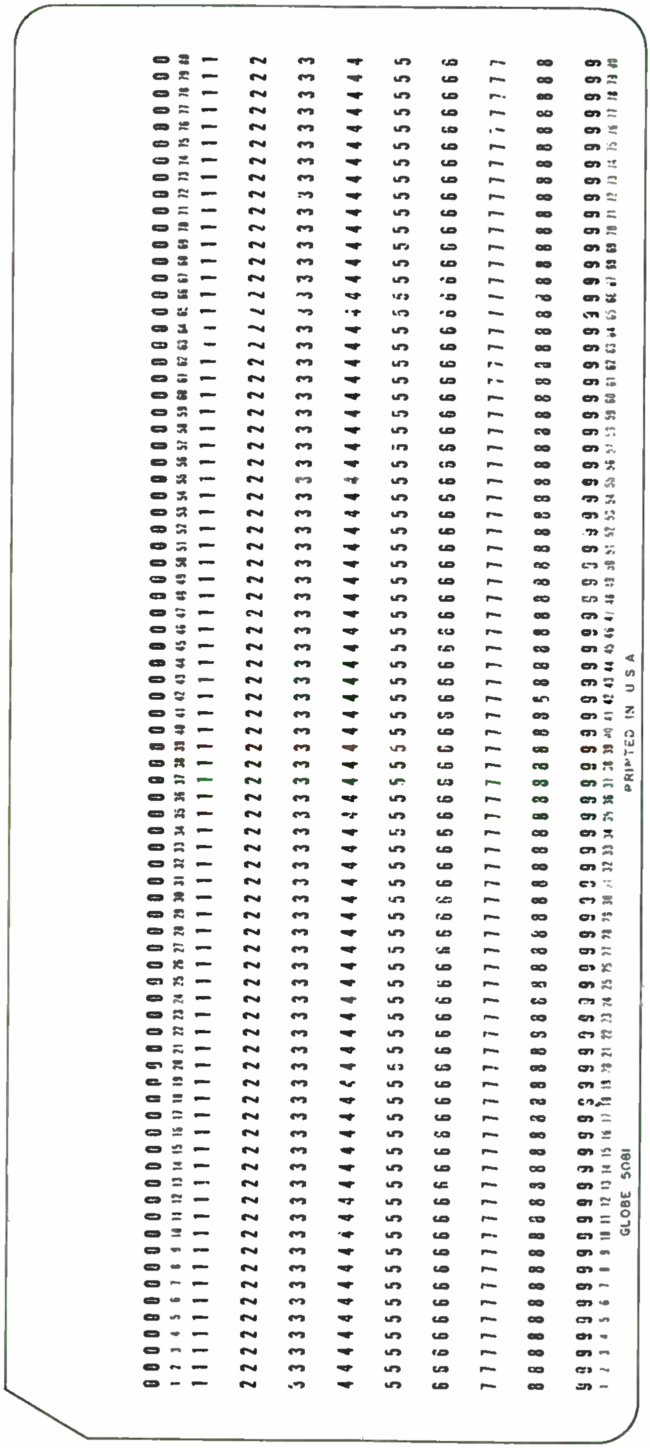
One of the major changes in the use of computers over the past decade has been in the way data are input. When computers first came into existence, data were input via computer cards. Advancements in technology have changed this: many computer centers now input information directly by means of a video terminal. However, computer card programming is not obsolete in many universities, colleges, and businesses. There are two reasons for this continued use: (1) replacing card equipment is very expensive; and (2) cards provide protection for the data—they cannot be purged (erased) by someone else, as can information that is input on a video terminal. Computer cards may be dated in terms of technological developments, but the fact remains that many researchers in both the public and the private sector still use them routinely.

Computer Cards

Computer cards are 3¼″ × 7⅜″ pieces of paper. Each contains 80 columns in which holes, representing numbers, letters, or characters, are punched by means of a keypunch machine. Each column on the card represents only a single data entry—one letter, character, or number. There are also 12 positions, or rows, on each card; the lower 10 positions correspond to the numbers 0 through 9, and the top 2 are used for characters or letters. A standard computer card is shown in Figure 16.1.

One of the major disadvantages of computer cards is that they must be handled carefully: any change in the card's original form ("do not fold, spindle, or mutilate") makes it impossible for the card reader

Figure 16.1 Standard computer card



to interpret the punched information. Cards can be ruined by damage that is invisible to the naked eye. Even exposure to humidity can render a card unreadable.

Computer cards containing numerous programs are generally fed into a computer simultaneously (or almost so). This process is referred to as a "batch job." The programs are then run one at a time (the order being determined by the amount of CPU time or central memory space each program requires).

The Keypunch Machine

A *keypunch machine* is similar to a standard typewriter and can be learned as easily. Unlike a typewriter, however, a keypunch machine perforates cards with patterns of holes that represent specific letters, characters, and numbers (which are also printed along the top of the card solely for the convenience of the researcher). Another difference between keypunch machines and typewriters is that the former use only capital letters; the shift key (known as the "numerical" key) is depressed to produce numbers and special characters.

Keypunch machines vary according to manufacturer, and specific details concerning their operation are not described here. The same basic procedure is followed with all models, however: blank cards are (1) inserted in the upper right-hand portion of the machine; (2) released individually into the central portion to be punched; and (3) stored in the upper left-hand portion of the machine.

Keypunching is made easier with the use of computer *coding sheets*, an example of which is shown in Figure 16.2. Each row on a coding sheet represents one computer card. Information to be punched is first written onto the sheets and then transferred to the cards. This eliminates the need to keypunch directly from survey questionnaires or other data-gathering instruments.

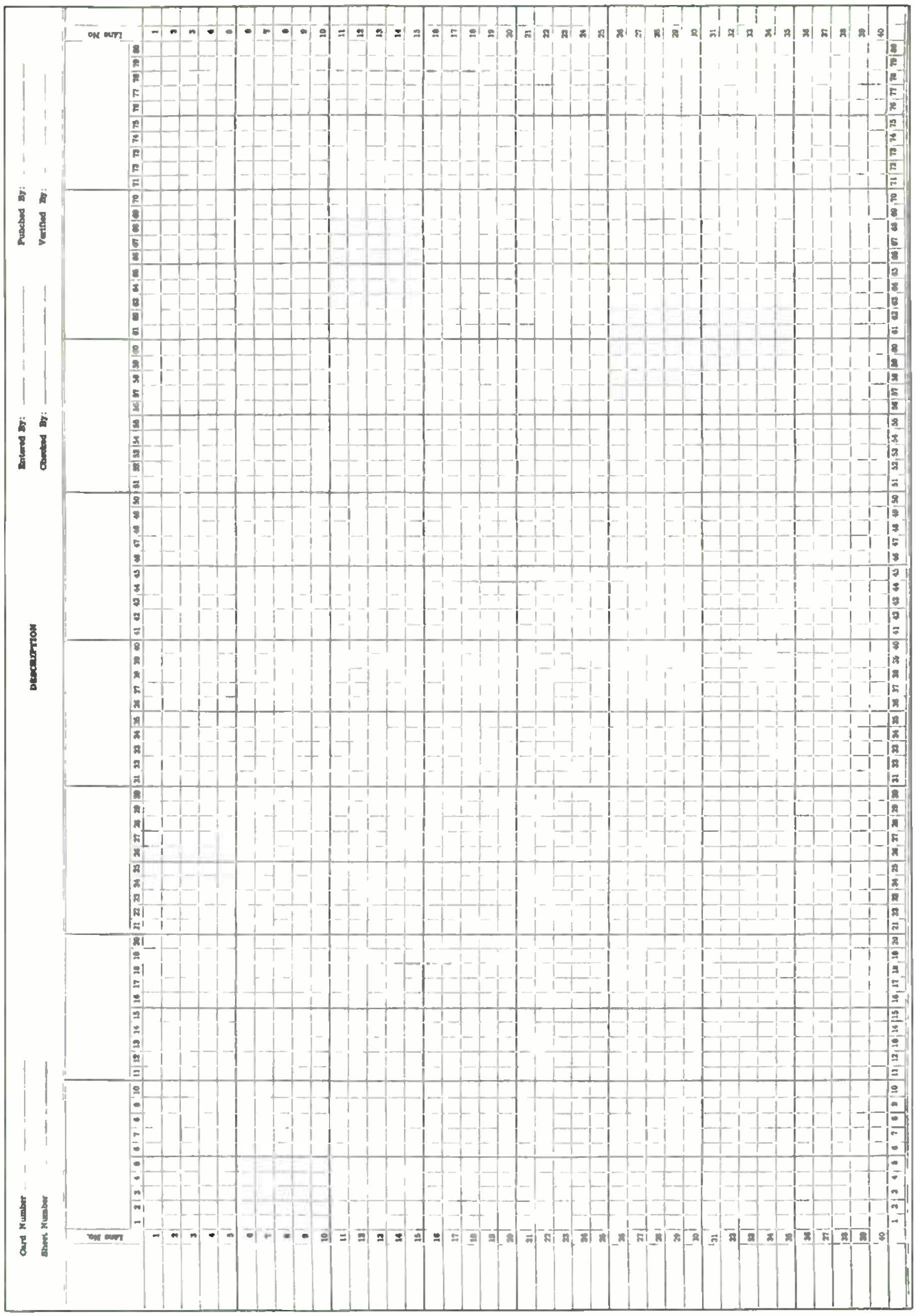
The Card Reader

A third piece of equipment used in computer card programming is the high-speed *card reader*, which electronically "reads" the holes punched in each card. The card reader is simple to operate: cards are placed in a loading area, and a key is pushed to begin the reading process. These machines are capable of reading several thousand cards per minute.

Magnetic Tapes and Disks

Magnetic tapes and disks are important in computer use, although researchers rarely come into direct contact with these items; they are still used primarily for external storage of data. Computer tapes are

Figure 16.2 Computer coding sheet



strips of plastic coated with a magnetic recording substance; they are similar to tapes used in audio and video recording. Each tape is capable of storing enough information to fill several thousand computer cards.

Disks are similar in appearance to phonograph records; they are also coated with a magnetic substance and have a high capacity for data storage. A disk is "read" by placing it on a *disk drive*, a high-speed machine that resembles a record player. The use of both disks and tapes is becoming more widespread; computer centers often find them an attractive alternative to cards because of their durability and their storage capacities.

Remote Terminals

Remote terminals provide access to a computer from a location outside of the center. A remote terminal may consist of a teletype machine, which operates exactly like a standard typewriter, or a video display terminal, which includes a keyboard and a small cathode ray tube screen.

The advantage of a remote terminal is that it can be placed wherever there is a telephone outlet; the receiver is held by a device called an **acoustical coupler**, which "calls" the central (or *main frame*) computer and maintains a connection over the phone lines. Researchers thus have at their disposal all of the capabilities of the main frame at a location convenient to them. Moreover, the recent boom in compact home computers has brought about an expansion of the remote access services offered by computer centers.

Using the Computer

The types of experience a person can acquire from using a computer are obviously dependent upon the facilities available to him or her. Most colleges and universities are equipped with some type of computer and offer students a chance to learn to use the equipment. However, freedom of access to computers varies among schools; some are more restrictive than others with regard to who is allowed into the center, what equipment they can use, and when they can use it. Some schools have very definite rules concerning when certain types of programs can be run. For example, programs which require a large amount of memory space or CPU time may be permitted only at night or only on certain days of the week.

Accordingly, the first step in using the computer is to learn the basic rules and procedures of the particular computer center in which it is located. This can be done by talking to people who have used the facilities before (instructors or other students) or to members of the computer center staff. An introduction to the facility is necessary to provide an acquaintance with the available equipment and how it is operated.

The second step for the beginning computer operator should be to select an easy program and use it to analyze a simple problem. Data for such an exercise are provided at the end of this chapter. It is important to read the manual for the selected program. The program manual explains the type and style of control cards to use, how to set up "decks" or sets of cards, and so on; it should be the first source consulted before using the computer, as well as whenever an error is encountered in the program.

Beginners should not worry about the quality of the analyses they conduct in early computer runs. The idea is to learn the basic procedures, not to produce significant results. One must start by analyzing relatively simple hypotheses in order to gain the experience necessary to tackle more meaningful and complex issues.

The third step is to use the computer on a consistent basis, in order to keep abreast of program changes and innovations. The SPSS program, for example, is constantly updated with releases of new subprograms, alterations to existing subprograms, and changes in all procedures. These can go unnoticed without regular use of the computer.

The following section provides examples from the SPSS program. They can be run on a computer using the raw data presented in Question 1 at the end of the chapter.

A Typical Computer Program

The SPSS program is one of the most widely used general utility packages in the United States. The program is relatively easy to learn and use, and it offers a variety of subprograms relevant to mass media research. The basics of the SPSS program can usually be acquired in a short period of time, since commands, control cards, and deck set-up instructions are worded in easy-to-understand language.

The data used in the following example were collected under the supervision of Dr. James Fletcher of the School of Journalism and Mass Communication at the University of Georgia. A subsample of Fletcher's original data has been selected, and some responses have been altered to provide a less complicated program. These data are presented solely to demonstrate the SPSS program and are not intended to reflect the scope of the original study. The SPSS runs included here were conducted by James B. Weaver, III, a graduate student at the University of Georgia.

The Fletcher study investigated the relationship between news source preference and public activity. That is, the study was intended to discover whether a person's level of public activity has any effect on his or her choice of news media: Do people active in public affairs prefer different sources of news than people who are less active?

The printout shown in Figure 16.3 demonstrates a simple FREQUENCIES run, which merely sums the variable responses and provides percentages. Each computer program has a unique printout. The one shown here is representative only of the SPSS program; other programs would produce similar results, but in a different format.

To facilitate discussion of the printout, each section of the program is identified by a capital letter. Each line (excluding those in Section G) represents the information on a single computer card. Notice how the various commands or control statements ("RUN NAME," "FILE NAME," "N OF CASES," and so on) are included in Columns 1–15; other information begins in Column 16. This is a unique feature of the SPSS program.

A	RUN NAME	This control card allows the researcher to name the program for sake of identification. This name is printed on all pages of the printout.
B	FILE NAME	This is an optional card whereby researchers can set up a temporary or permanent file for the program. The card here indicates that the data will be stored on tape for use at a later time.
C	VARIABLE LIST	This card requires the researcher to provide a list of variables to be included in the analysis. Since the names are limited to eight characters, it is necessary on occasion to use abbreviations, such as "RSPDNT" for *respondent.* Variables can simply be numbered to save keypunch strokes.
D	INPUT MEDIUM	This card describes the method by which the data will be entered into the computer: tape, disk, or card.
E	N OF CASES	This optional card indicates the number of subjects involved in the analysis.
F	INPUT FORMAT	This control card tells the computer how to read the data cards. The format indicates which columns contain relevant information and which columns are to be skipped. In order to make verification of cards easier, it is suggested that blank columns be left between variables or sets of variables (as has been done here) so

Figure 16.3 SPSS frequencies subprogram

```
A  RUN NAME          NEWS PREFERENCE VS. PUBLIC ACTIVITIES SCALE
B  FILE NAME         ATHENS NEWS MARKET
C  VARIABLE LIST     RSPNDNT,NEWSPREF,RADPREF,TVPREF,MAGPREF,ISSUE,
                     LTR,MTG,LOCAL,CAMPGN,PASCALE,AGE,SEX,DAY,TIME
D  INPUT MEDIUM      CARD
E  N OF CASES        100
F  INPUT FORMAT      FIXED(F3.0,1X,F1.0,1X,F2.0,1X,2F1.0,1X,6F1.0,1X,F2.0,1X,
                     F1.0,1X,F2.0,F4.0)

   ACCORDING TO
   YOUR INPUT
   FORMAT,
   VARIABLES ARE TO
   BE READ AS
   FOLLOWS

         VARIABLE    FORMAT    RECORD    COLUMNS

         RSPNDNT     F 3.0       1         1-3
         NEWSPREF    F 1.0       1         5-5
         RADPREF     F 2.0       1         7-8
         TVPREF      F 1.0       1        10-10
         MAGPREF     F 1.0       1        11-11
         ISSUE       F 1.0       1        13-13
G        LTR         F 1.0       1        14-14
         MTG         F 1.0       1        15-15
         LOCAL       F 1.0       1        16-16
         CAMPGN      F 1.0       1        17-17
         PASCALE     F 1.0       1        18-18
         AGE         F 2.0       1        20-21
         SEX         F 1.0       1        23-23
         DAY         F 2.0       1        25-26
         TIME        F 4.0       1        27-30

   THE INPUT FORMAT PROVIDES FOR 15 VARIABLES. 15 WILL BE READ. IT
   PROVIDES FOR 1 RECORDS ("CARDS") PER CASE. A MAXIMUM OF 30
   "COLUMNS" ARE USED ON A RECORD.

H  VAR LABELS        RSPNDNT,RESPONDENT NUMBER/NEWSPREF,
                     NEWSPAPER STATION PREFERRED/RADPREF,RADIO
                     STATION PREFERRED/TVPREF, TV NEWS PREFERRED/
                     MAGPREF, NEWSMAGAZINE PREFERRED/ISSUE,
                     INVOLVED IN PUBLIC ISSUES/LTR,WRITTEN TO EDITOR/
                     MTG, ADDRESSED A MEETING/LOCAL,SUPPORTED A
                     LOCAL ISSUE/CAMPGN, ACTIVE IN POLITICAL CAMPAIGN/
                     PASCALE, PUBLIC ACT. SCALE
I  VALUE LABELS      NEWSPREF (0) NONE (1) DAILY NEWS (2) BANNER HERALD
                     (3) OBSERVER (4) RED AND BLACK/RADPREF (00) NONE
                     (01) WUOGFM (02) WGAU (03) WNGCFM (04) WRFC (05)
                     WAGQFM (06) WFOXFM (07) WCCD (08) OTHER/TVPREF (0)
                     NONE (1) WSBTV (2) WAGATV (3) WXIATV/MAGPREF (0)
                     NONE (1) TIME (2) NEWSWEEK (3) US NEWS (4) OTHER/SEX
                     (0) FEMALE (1) MALE/ISSUE (0) NO (1) YES/LTR (0) NO (1)
                     YES/MTG (0) NO (1) YES/LOCAL (0) NO (1) YES/CAMPGN (0)
                     NO (1) YES/
J  PRINT FORMATS     PASCALE TO AGE(3)
K  FREQUENCIES       GENERAL = NEWSPREF TO SEX
L  STATISTICS        ALL
M  OPTIONS           8,9
N  READ INPUT DATA
```

	that errors can be spotted more easily. The coded sequence listed here is interpreted in the following manner: An "F" represents a numeric variable; an alphanumeric variable would be coded as "A" (other symbols are also used). The number preceding the "F" tells the computer how many variables to read, and the number following it indicates the number of columns for each variable. Thus the statement "2F1.0" instructs the computer to read two single-column numerical variables. The command "1X" means to skip one column.
G	This section is a convenience provided by the computer in the SPSS program; it displays the information given by the operator on the "INPUT FORMAT" card. Thus researchers can know immediately if an incorrect format statement has been entered.
H VAR LABELS	This is another optional card; it allows the researcher to expand the variable names listed on the "VARIABLE LIST" card and instructs the computer to use these full names on the final printout.
I VALUE LABELS	This optional card assigns names and values to the possible responses for each variable in the analysis. For example, "1" may represent a yes answer and "0" a no answer to a specific question.
J PRINT FORMATS	This card indicates how many digits to the right of the decimal point should be printed for specific variables. In the example, the computer will calculate each value to three decimal places for variables "PASCALE" through "AGE."
K FREQUENCIES	This card instructs the computer to determine the statistical frequencies of the variables named. In this instance, frequency distri-

		butions for 12 variables ("NEWSPREF" through "SEX") will be computed.
L	STATISTICS	This card is used to request specific statistical procedures for the FREQUENCIES subprogram; "ALL" requests that every statistic available be computed.
M	OPTIONS	The options card allows the researcher to request additional services from the program. In the example, Option 8 requests histograms for each variable; Option 9 requests a table of contents for each variable.
N	READ INPUT DATA	This indicates that the data cards will follow immediately.

Examples from the printout of the FREQUENCIES subprogram are shown in Figures 16.4 and 16.5. The frequency display for the variable "RADPREF" (radio station preferred) is shown in 16.4.

Figure 16.5 shows the same variable in histogram form (Option 8), providing an alternate visual display of the data.

Another frequently used subprogram of the SPSS package is

Figure 16.4 RADPREF frequency distribution (Table)

NEWS PREFERENCE VS. PUBLIC ACTIVITIES
RADPREF RADIO STATION PREFERRED

CATEGORY LABEL	CODE	ABSOLUTE FREQ	RELATIVE FREQ (PCT)	ADJUSTED FREQ (PCT)	CUM FREQ (PCT)
NONE	0	19	19.0	19.0	19.0
WUOGFM	1.	6	6.0	6.0	25.0
WGAU	2.	8	8.0	8.0	33.0
WNGCFM	3.	12	12.0	12.0	45.0
WRFC	4.	8	8.0	8.0	53.0
WAGQFM	5.	30	30.0	30.0	83.0
WFOXFM	6.	5	5.0	5.0	88.0
OTHER	8.	12	12.0	12.0	100.0
	TOTAL	100	100.0	100.0	

Figure 16.5 RADPREF frequency distribution (histogram)

```
NEWS PREFERENCE VS. PUBLIC ACTIVITIES
RADPREF        RADIO STATION PREFERRED


      CODE

           I
        0  ******************* (19)
           I NONE
           I
        1. ****** (6)
           I WUOGFM
           I
        2. ******** (8)
           I WGAU
           I
        3. ************ (12)
           I WNGCFM
           I
        4. ******** (8)
           I WRFC
           I
        5. ****************************** (30)
           I WAGQFM
           I
        6. ***** (5)
           I WFOXFM
           I
        8. ************ (12)
           I OTHER
           I
           I.........I.........I.........I.........I.........I
           0        10        20        30        40        50

           FREQUENCY

MEAN          3.660      STD ERR      .253      MEDIAN      4.125

MODE          5.000      STD DEV     2.528      VARIANCE    6.388

KURTOSIS      -.916      SKEWNESS     .021      RANGE       8.000

MINIMUM           0      MAXIMUM     8.000      SUM       366.000

C.V. PCT 69.058          .95 C.I.    3.158

VALID CASES 100          MISSING CASES 0
```

CROSSTABS, which performs the chi-square contingency table analysis described in Chapter 11. This subprogram offers researchers a useful and efficient tool to investigate mass media questions. An example of an SPSS CROSSTABS printout is shown in Figure 16.6, also using the Fletcher data. As can be seen from the printout, only a few of the control cards are different from those used in the FREQUENCIES run (the "FREQUENCIES"/"CROSSTABS," "STATISTICS," and "OP-

Figure 16.6 SPSS crosstabs subprogram

```
RUN NAME          NEWS PREFERENCE VS. PUBLIC ACTIVITIES SCALE
FILE NAME         ATHENS NEWS MARKET
VARIABLE LIST     RSPNONT, NEWSPREF,RAOPREF,TVPREF,MAGPREF,ISSUE,LTR,
                  MTG,LOCAL,CAMPGN,PASCALE,AGE,SEX,OAY,TIME
INPUT MEOIUM      CARO
N OF CASES        100
INPUT FORMAT      FIXEO(F3.0,1X,F1.0,1X,F2.0,1X,2F1.0,1X,6F1.0,1X,F2.0,1X,
                  F1.0,1X,F2.0,F4.0)
ACCOROING TO YOUR INPUT FORMAT, VARIABLES ARE TO BE REAO AS FOLLOWS

     VARIABLE          FORMAT          RECORO          COLUMNS

     RSPNONT           F 3.0             1               1-3
     NEWSPREF          F 1.0             1               5-5
     RAOPREF           F 2.0             1               7-B
     TVPREF            F 1.0             1              10-10
     MAGPREF           F 1.0             1              11-11
     ISSUE             F 1.0             1              13-13
     LTR               F 1.0             1              14-14
     MTG               F 1.0             1              15-15
     LOCAL             F 1.0             1              16-16
     CAMPGN            F 1.0             1              17-17
     PASCALE           F 1.0             1              1B-1B
     AGE               F 2.0             1              20-21
     SEX               F 1.0             1              23-23
     OAY               F 2.0             1              25-26
     TIME              F 4.0             1              27-30

THE INPUT FORMAT PROVIOES FOR 15 VARIABLES. 15 WILL BE REAO.
IT PROVIOES FOR 1 RECOROS (*CAROS*) PER CASE.
A MAXIMUM OF 30 *COLUMNS* ARE USEO ON A RECORO

VAR LABELS        RSPNONT,RESPONOENT NUMBER/NEWSPREF,NEWPAPER
                  PREFERREO/RAOPREF, RAOIO STATION PREFERREO/TVPREF, TV
                  NEWS PREFERREO/MAGPREF, NEWS MAGAZINE PREFERREO/
                  ISSUE, INVOLVEO IN PUBLIC ISSUES/LTR, WRITTEN TO EOITOR/
                  MTG, AOORESSEO A MEETING/LOCAL, SUPPORTEO A LOCAL
                  ISSUE/CAMPGN, ACTIVE IN POLITICAL CAMPAIGN/PASCALE,
                  PUBLIC ACTIVITIES SCALE
VALUE LABELS      NEWSPREF(0)NONE(1) OAILY NEWS(2)BANNER
                  HERALO(3)OBSERVER (4)REO AND BLACK/
                  RAOPREF(00)NONE(01)WUOGFM(02)WGAU(03)WNGCFM(04)
                  WRFC(05)WAGQFM(06)WFOXFM(07)WCCO(0B)OTHER/TVPREF(0)
                  NONE(1)WSBTV(2)WAGATV(3)WXIATV/MAGPREF(0)NONE(1)TIME
                  (2)NEWSWEEK(3)USNEWS(4)OTHER/SEX(0)FEMALE(1)MALE/
                  ISSUE(0)NO(1)YES/LTR(0)NO(1)YES/MTG(0)NO(1)YES/LOCAL(0)
                  NO(1)YES/CAMPGN(0)NO(1)YES/
PRINT FORMATS     PASCALE TO AGE(3)
CROSSTABS         TABLES=RAOPREF BY NEWSPREF,TVPREF,MAGPREF,PASCALE,
                  SEX
STATISTICS        ALL
OPTIONS           9
REAO INPUT OATA
```

TIONS" cards). A CROSSTABS printout comparing the variables "RADPREF" (radio preference) and "NEWSPREF" (newspaper preference) is shown in Figure 16.7.

Figure 16.7 CROSSTABS example

* * * * * * * * * * * * C R O S S T A B U L A T I O N O F * * * * * * * * * * *

RADPREF RADIO STATION PREFERRED
BY NEWSPREF NEWSPAPER PREFERRED

* *

RADPREF	NEWSPREF: COUNT / ROW PCT / COL PCT / TOT PCT	NONE 0.	DAILY NEWS 1.	BANNER HERALD 2.	OBSERVER 3.	RED AND BLACK 4.	ROW TOTAL
0. NONE	COUNT	4	1	10	1	3	19
	ROW PCT	21.1	5.3	52.6	5.3	15.8	19.0
	COL PCT	28.6	9.1	20.0	10.0	20.0	
	TOT PCT	4.0	1.0	10.0	1.0	3.0	
1. WUOGFM	COUNT	2	0	2	1	1	6
	ROW PCT	33.3	0	33.3	16.7	16.7	6.0
	COL PCT	14.3	0	4.0	10.0	6.7	
	TOT PCT	2.0	0	2.0	1.0	1.0	
2. WGAU	COUNT	2	1	4	1	0	8
	ROW PCT	25.0	12.5	50.0	12.5	0	6.0
	COL PCT	14.3	9.1	8.0	10.0	0	
	TOT PCT	2.0	1.0	4.0	1.0	0	
3. WNGCFM	COUNT	0	3	8	1	0	12
	ROW PCT	0	25.0	66.7	8.3	0	12.0
	COL PCT	0	27.3	16.0	10.0	0	
	TOT PCT	0	3.0	8.0	1.0	0	
4. WRFC	COUNT	1	0	5	1	1	8
	ROW PCT	12.5	0	62.5	12.5	12.5	8.0
	COL PCT	7.1	0	10.0	10.0	6.7	
	TOT PCT	1.0	0	5.0	1.0	1.0	
5. WAGQFM	COUNT	4	4	12	3	7	30
	ROW PCT	13.3	13.3	40.0	10.0	23.3	30.0
	COL PCT	28.6	36.4	24.0	30.0	46.7	
	TOT PCT	4.0	4.0	12.0	3.0	7.0	
6. WFOXFM	COUNT	0	0	3	0	2	5
	ROW PCT	0	0	60.0	0	40.0	5.0
	COL PCT	0	0	6.0	0	13.3	
	TOT PCT	0	0	3.0	0	2.0	
8. OTHER	COUNT	1	2	6	2	1	12
	ROW PCT	8.3	16.7	50.0	16.7	8.3	12.0
	COL PCT	7.1	18.2	12.0	20.0	6.7	
	TOT PCT	1.0	2.0	6.0	2.0	1.0	
COLUMN TOTAL		14	11	50	10	15	100
		14.0	11.0	50.0	10.0	15.0	100.0

Further Suggestions for Operators

Experience is important in using any type of electronic equipment, and computers are no exception. One can read books and listen to lectures about how to operate computers, but there is no substitute for actually sitting down and using one. Researchers who use the computer on a consistent basis quickly learn shortcuts for many procedures. They also learn to avoid minor problems that can arise in, for example, keypunching or entering specific commands. Below is a list of suggestions that may facilitate computer use. They can help the beginning operator to deal with some of the basic problems he or she may encounter.

1. Read the program manual for any computer package before attempting to run a program. In many cases it is not necessary to read the entire manual, only those sections pertaining to a specific subprogram such as CROSSTABS. The manual is an essential guide to using any canned computer program.

2. Never assume that the data have been keypunched or typed correctly; all cards and entries should be double- or triple-checked for accuracy. Data input errors are common, and a great deal of time and money can be saved if the data are proofread before the program is run. One excellent way to verify data involves the cooperation of two people: one person reads the coding sheet data aloud while the other follows along on the punched cards or CRT screen.

3. When computer cards are used, number all of the cards in the deck with a pencil—a small number in the upper right hand corner will suffice. This will save time and a good deal of frustration if the deck should accidentally be dropped.

4. If a set of data cards will be used several times, it is suggested that the data be stored on tape or disk. The method for such an operation can be explained by a local computer operator. Continual use of cards creates wear and tear that may eventually necessitate duplicating the entire deck.

5. Ask questions. Computer operators, programmers, and users usually share a "scientific camaraderie" that is noticeable in most computer centers. People who use the computer seem to belong to a "club" whose members are anxious to help one another, so do not be afraid to ask for assistance.

6. Try several different computer packages to determine which works best in different situations. Some packages, such as SPSS, have user's manuals that are easy to follow and are excellent for beginners; other packages, such as BMD-P, have more complicated manuals that may require the assistance of someone who has used the package. Every researcher has a favorite computer package, but it is probably unwise to stay with one package for every type of analysis. Some may

be excellent for multivariate statistics, for example, while others are preferable for univariate procedures. Additionally, some packages (such as SAS) can be used only with certain types of main frame computers.

Thus there are many factors to consider when choosing a package. If more than one program seems appropriate for a particular study, the researcher should examine each by asking: Is the user's manual well documented and easy to follow? Are simple instructions provided for deck set-up? How many different statistical methods does the package offer? Is the printout of good quality and easy to interpret? Can the data be easily stored and manipulated? Beginning computer users would do well to select a single all-purpose package and use it at first for running all programs. The authors personally recommend the SPSS package as being the most versatile and easy-to-use package available at this time. In addition, once a researcher is familiar with the use of SPSS, learning other packages is much simpler.

Summary

Mass media researchers regularly use the computer in all types of analyses. Once considered a unique research tool, the computer is now regarded as essential in almost every mass media study. This chapter introduces computers and explains the basic procedures involved in their use. Readers are strongly encouraged to pursue the suggested readings listed at the end of the chapter for more information about computers and particular computer packages.

Questions and Problems for Further Investigation

1. The raw data for the SPSS example provided in this chapter are presented below. The "VAR LABELS" and "VALUE LABELS" control cards, as shown in Figures 16.3 and 16.6, define the data. Compute frequencies and crosstabs for any of the variables listed below. What are some other types of statistical analyses for which these data can be used?

```
RUN NAME        NEWS PREFERENCE VS. PUBLIC ACTIVITIES SCALE
FILE NAME       ATHENS NEWS MARKET
VARIABLE LIST   RSPNDNT,NEWSPREF,RADPREF,TVPREF,MAGPREF,ISSUE,LTR,
                MTG,LOCAL,CAMPGN,PASCALE,AGE,SEX,DAY,TIME
INPUT MEDIUM    CARD
N OF CASES      100
INPUT FORMAT    FIXED (F.3,1X,F1.0,1X,F2.0,1X,2F1.0,1X,6F1.0,1X,F2.0,1X,F1.0,
                1X,F2.0,F4.0)
VAR LABELS      RSPNDNT,RESPONDENT NUMBER/NEWSPREF,NEWSPAPER
                PREFERRED/RADPREF RADIO STATION PREFERRED/TVPREF,
                TV NEWS PREFERRED/MAGPREF, NEWS MAGAZINE
                PREFERRED/ISSUE, INVOLVED IN PUBLIC ISSUES/LTR WRITTEN
                TO EDITOR/MTG, ADDRESSED A MEETING/LOCAL, SUPPORTED
                LOCAL ISSUE/CAMPGN, ACTIVE IN POLITICAL CAMPAIGN/
                PASCALE, PUB ACTIVITIES SCALE
```

```
VALUE LABELS    NEWSPREF (0)NONE(1)DAILY NEWS (2) BANNER HERALD (3)
                OBSERVER (4) RED AND BLACK/
                RADPREF (00) NONE (01) WUDGFM (02) WGAU (03) WNGCFM (04)
                WRFC (05) WAGQFM (06) WFOXFM (07) WCCD (08) OTHER/
                TVPREF (0) NONE (1) WSBTV (2) WAGATV (3) WXIATV/
                MAGPREF (0) NONE (1) TIME (2) NEWSWEEK (3) U S NEWS (4)
                OTHER/
                SEX (0) FEMALE (1) MALE/
                ISSUE (0) NO (1) YES/
                LTR   (0) NO (1) YES/
                MTG   (0) NO   (1) YES/
                LOCAL (0) NO   (1) YES/
                CAMPGN (0) NO  (1) YES/
PRINT FORMATS   PASCALE TO AGE(3)
CROSSTABS       TABLES = RADPREF BY NEWSPREF, TVPREF
STATISTICS      ALL
OPTION          9
READ INPUT DATA
001 3 08 24 011103 23 0 141550
002 2 08 34 000000 21 0 141553
003 0 01 02 011103 26 0 141556
004 4 05 11 100113 22 0 141606
005 2 05 20 000101 20 1 141613
006 4 05 12 000000 20 1 141616
007 0 00 20 011002 27 0 141625
008 2 06 32 001113 17 0 141628
009 2 00 11 111114 30 0 141624
010 2 08 32 001102 20 1 141632
011 2 00 20 000000 50 1 141635
012 4 05 01 000011 20 1 141638
013 4 00 00 010001 26 1 141646
014 2 02 30 000000 48 1 142037
015 2 02 34 000000 73 1 142043
016 1 05 20 001001 35 1 142048
017 2 00 02 010113 34 1 142050
018 4 05 20 101103 21 0 114211
019 2 06 10 000000 31 1 114213
020 2 08 30 010001 16 1 142152
021 0 02 24 001001 70 1 151800
022 2 08 20 000000 76 0 151810
023 3 05 32 001113 24 0 151820
024 2 05 20 001001 20 0 151830
025 2 05 30 000000 20 0 151840
026 2 03 10 000000 50 0 151850
027 1 03 11 000000 38 1 151900
028 2 05 24 000000 25 0 151905
029 2 00 20 000000 77 0 151910
030 3 03 01 000112 34 0 151915
031 4 06 21 110002 27 1 151920

RUN NAME        NEWS PREFERENCE VS. PUBLIC ACTIVITIES SCALE
FILE NAME       ATHENS NEWS MARKET
VARIABLE LIST   RSPNDNT, NEWSPREF, RADPREF, TVPREF, MAGPREF, ISSUE, LTR,
                MTG, LOCAL, CAMPGN, PASCALE, AGE, SEX, DAY, TIME
INPUT MEDIUM    CARD
N OF CASES      100
INPUT FORMAT    FIXED (F3.0,1X,F1.0,1X,F2.0,1X,2F1.0,1X,6F1.0,1X,F2.0,1X,
                F1.0,1X,F2.0,F4.0)
VAR LABELS      RSPNDNT,RESPONDENT NUMBER/NEWSPREF,NEWSPAPER
                PREFERRED/RADPREF, RADIO STATION PREFERRED/TVPREF,
                TV NEWS PREFERRED/MAGPREF, NEWS MAGAZINE
                PREFERRED/ISSUE, INVOLVED IN PUBLIC ISSUES/LTR, WRITTEN
                TO EDITOR/MTG, ADDRESSED A MEETING/LOCAL, SUPPORTED A
                LOCAL ISSUE/CAMPGN, ACTIVE IN POLITICAL CAMPAIGN/
                PASCALE, PUBLIC ACTIVITIES SCALE
```

```
VALUE LABELS     NEWSPREF (0) NONE (1) OAILY NEWS (2) BANNER HERALO (3)
                 OBSERVER (4)REO AND BLACK/
                 RADPREF (00) NONE (01) WUOGFM (02) WGAU (03) WNGCFM (04)
                 WRFC (05) WAGQFM (06) WFOXFM (07) WCCO (08) OTHER/
                 TVPREF (0) NONE (1) WSBTV (2) WAGATV (3) WXIATV/
                 MAGPREF (0) NONE (1) TIME (2) NEWSWEEK (3) U S NEWS (4)
                 OTHER/
                 SEX (0) FEMALE (1) MALE/
                 ISSUE (0) NO (1) YES/
                 LTR (0) NO (1) YES/
                 MTG (0) NO (1) YES/
                 LOCAL (0) NO (1) YES/
                 CAMPGN (0) NO (1) YES/
PRINT  FORMATS   PASCALE TO AGE (3)
CROSSTABS        TABLES=RAOPREF BY NEWSPREF, TVPREF, MAGPREF, PASCALE,
                 SEX
STATISTICS       ALL
OPTION           9
REAO INPUT OATA
001 3 08 24 011103 23 0 141550
002 2 08 34 000000 21 0 141553
003 0 01 02 011103 26 0 141556
004 4 05 11 100113 22 0 141606
005 2 05 20 000101 20 1 141613
006 4 05 12 000000 20 1 141616
007 0 00 20 011002 27 0 141625
008 2 06 32 001113 17 0 141628
009 2 00 11 111114 30 0 141624
010 2 08 32 001102 20 1 141632
011 2 00 20 000000 50 1 141635
012 4 05 01 000011 20 1 141638
013 4 00 00 010001 26 1 141646
014 2 02 30 000000 48 1 142037
015 2 02 34 000000 73 1 142043
016 1 05 20 001001 35 1 142048
017 2 00 02 010113 34 1 142050
018 4 05 20 101103 21 0 114211
019 2 06 10 000000 31 1 114213
020 2 08 30 010001 16 1 142152
021 0 02 24 001001 70 1 151800
022 2 08 20 000000 76 0 151810
023 3 05 32 001113 24 0 151820
024 2 05 20 001001 20 0 151830
025 2 05 30 000000 20 0 151840
026 2 03 10 000000 50 0 151850
027 1 03 11 000000 38 1 151900
028 2 05 24 000000 25 0 151905
029 2 00 20 000000 77 0 151910
030 3 03 01 000112 34 0 151915
031 4 06 21 110002 27 1 151920
032 2 05 21 111114 30 0 151925
033 2 01 10 000000 23 1 151930
034 2 00 10 100001 45 0 151935
035 2 03 24 000011 50 1 151940
036 4 00 00 000000 28 1 151945
037 3 01 21 011114 25 1 151955
038 1 03 22 100001 54 1 152000
039 2 01 20 000101 28 0 152010
040 3 08 10 001001 52 0 152020
041 0 00 00 000000 80 0 151415
042 1 08 30 010001 33 0 151425
043 2 05 34 101114 17 1 151429
044 2 03 14 000000 23 1 151431
045 2 05 21 000101 16 0 151435
```

```
046 4 00 30 000000 31 0 151437
047 4 05 00 111104 19 0 151440
048 4 01 12 001001 22 1 151445
049 2 05 10 001000 21 1 151450
050 1 08 10 011114 72 0 151452
051 1 05 20 000101 16 0 151500
052 0 08 10 000011 27 0 151801
053 4 05 30 111114 20 0 151505
054 2 04 30 000000 33 1 151510
055 3 05 00 000000 27 0 151737
056 0 00 10 000000 59 0 151741
057 2 00 20 001001 11 1 151747
058 0 01 30 100102 28 1 151750
059 0 05 12 000000 27 1 151752
060 2 08 11 010102 44 0 151755
061 0 05 30 000000 26 0 151900
062 3 04 12 001113 22 1 151915
063 2 05 20 000000 22 1 151927
064 4 04 21 000000 30 1 151940
065 0 00 30 001113 00 0 151958
066 1 05 24 000000 28 1 152000
067 3 02 30 000000 23 0 152052
068 4 06 34 001001 21 0 152075
069 2 04 14 000000 23 1 152090
070 2 00 00 000000 23 0 161040
071 3 00 11 011013 25 1 161065
072 2 02 04 001012 59 1 161075
073 0 05 10 000000 20 0 161210
074 1 03 30 000000 31 0 161225
075 0 04 34 000000 62 0 161235
076 2 00 14 011002 72 0 161243
077 2 05 02 000112 55 0 161257
078 2 03 24 000000 10 1 201930
079 0 05 30 000000 24 1 201940
080 2 05 21 011002 33 1 201955
081 2 02 20 111104 33 0 151908
082 1 02 21 100012 66 0 151916
083 2 04 20 000000 49 0 151921
084 2 04 30 000000 28 1 151927
085 2 00 13 000000 60 0 151935
086 4 05 12 001001 19 0 151938
087 3 05 20 110103 19 0 151943
088 2 04 20 010001 52 1 151945
089 2 00 20 000000 61 1 151951
090 2 06 20 011103 31 1 151955
091 2 03 20 000000 34 0 151958
092 2 03 20 000000 73 0 152001
093 1 00 20 101002 53 1 152014
094 0 02 20 000000 82 0 152026
095 1 05 32 010001 52 0 152031
096 2 08 31 010012 48 0 152100
097 2 03 30 010001 53 0 152035
098 4 08 10 101103 68 1 152038
099 2 03 10 111104 33 0 152047
100 2 05 10 001001 27 1 152127
FINISH
```

2. The BMD-P package is often used by mass media research for certain types of problems. The BMD-P control cards listed below contain the same data as in the SPSS example and are set up for running a frequencies program. They are provided for the benefit of readers who might wish to compare the printouts generated by the two programs.

PROGRAM CONTROL INFORMATION

```
/PROBLEM   TITLE IS 'NEWS PREFERENCE VS. PUBLIC ACTIVITIES SCALE'.
/INPUT     VARIABLES ARE 15.
           FORMAT IS '(F3.0,1X,F1.0,1X,F2.0,1X,2F1.0,1X,6F1.0,1X,F2.0,
           1X,F1.0,1X,F2.0,F4.0)'.
           CASES ARE 100.
/VARIABLE  NAMES ARE RSPNDNT,NEWSPREF,RADPREF,TVPREF,MAGPREF,
           ISSUE,LTR,MTG,LOCAL,CAMPGN,PASCALE,AGE,SEX,DAY,TIME.
/TABLE     COLUMN IS RADPREF.
           ROW IS  NEWSPREF,TVPREF,MAGPREF,PASCALE,SEX.
           CROSS.
/GROUPS    CODES(2)  ARE  0,1,2,3,4.
           NAMES(2)  ARE  NONE,DAILYNEWS,BANNER,OBSERVER,
           REDBLACK.
           CODES(3)  ARE  00,01,02,03,04,05,06,07,08.
           NAMES(3)  ARE  NONE,WUOGFM,WGAU,WNGCFM,WRFC,
           WAGQFM,WFOXFM,WCCD,OTHER.
           CODES(4)  ARE  0,1,2,3.
           NAMES(4)  ARE  NONE,WSBTV,WAGATV,WXIATV.
           CODES(5)  ARE  0,1,2,3,4.
           NAMES(5)  ARE  NONE,TIMEMAG,NEWSWEEK,USNEWS,OTHER.
           CODES(6)  ARE  0,1.
           NAMES(6)  ARE  NO,YES.
           CODES(7)  ARE  0,1.
           NAMES(7)  ARE  NO,YES.
           CODES(8)  ARE  0,1.
           NAMES(8)  ARE  NO,YES.
           CODES(9)  ARE  0,1.
           NAMES(9)  ARE  NO,YES.
           CODES(10)  ARE  0,1.
           NAMES(10)  ARE  NO,YES.
           CODES(11)  ARE  0,1,2,3,4.
           NAMES(11)  ARE  NOT,  LOW,  '-',  '+',  HIGH.
           CUTPOINTS(12)  ARE  0,25,35,45,55,65.
           NAMES(12)  ARE  REFUSED,  '25ORLESS','26  TO  35','36  TO  45',
           '46
           TO  55','56  TO  65',"OVER  65'.
           CODES(13)  ARE  0,1.
           NAMES(13)  ARE  FEMALE,MALE.
/END
```

References and Suggested Readings

Barr, A. J., Goodnight, J. H., Sall, J. R., and Helwig, J. T. 1976. *SAS: Statistical Analysis System*. Raleigh, N. C.: SAS Institute Inc.

Brier, A., and Robinson, J. 1974. *Computers and the Social Sciences*. New York: Columbia University Press.

Dixon, W., ed. 1973. *Biomedical Computer Programs*. Berkeley: University of California Press.

Finn, J. D. 1978. *Multivariance: Univariate and Multivariate Analysis of Variance, Covariance, and Regression*. Chicago: National Educational Resources.

Inter-University Consortium for Political Research. 1973. *OSIRIS III*. Ann Arbor: University of Michigan.

Klecka, W. R., Nie, N. H., and Hull, C. H. 1975. *SPSS Primer*. New York: McGraw-Hill.

Nie, N. H., Bent, D., and Hull, C. H. 1975. *Statistical Package for the Social Sciences.* New York: McGraw-Hill.

Rattenbury, J., and Pelletier, P. 1974. *Data Processing in the Social Sciences with OSIRIS.* Ann Arbor: University of Michigan.

CHAPTER 17

Research Reporting, Ethics, and Financial Support

The previous chapters have discussed mass media research from the initial planning stages of a study to the selection of the most appropriate statistical methods for testing purposes. This chapter focuses on three other areas that are not part of the research process itself but are nevertheless vital to the execution of any research project: reporting, ethics, and financial support.

Research Reports

Mass media researchers routinely encounter three types of research reports: those intended for publication in professional journals, those designed for conventions or professional meetings, and those prepared for management use in decision-making. The latter type of report tends to be less specific in describing the methodologies used, since management does not generally require such information. However, this does not mean that private sector research reports can completely eliminate any important information.

The Need for Accurate Reporting Procedures

There are two reasons why researchers need to report research accurately. The first is that a clear explanation of the methods involved provides an opportunity for readers to more completely understand the research project. Researchers should keep in mind that, in most cases, a reader's knowledge of their project is based solely on the information contained in their report. Since readers do not instinctively understand each procedure used in a study, these details must be provided. Second, an accurate report provides the necessary information for those who wish to replicate the study. As Rummel (1970) suggested:

> Enough information must be included or filed somewhere in public archives to enable reproduction of the study without the necessity of personal contact with the investigator. This is to ensure that a study is always replicable regardless of the decades or generations that may pass.

Rummel has even argued that researchers should be able to replicate a published study from the information contained therein. Realistically speaking, however, this may not be possible in all cases. Mass media journals have limited space, and journal editors do not have the luxury of including all raw data, tables, and graphs involved in a study; they are forced to eliminate some essential information. Therefore, Rummel's alternative—data archives—is very important. Unfortunately, the mass media field has yet to establish its own data archive service for researchers to use.

The conclusion, then, is that individual researchers must take full responsibility for accurately reporting and storing their own research data. To facilitate this task, the following subsections describe the important elements of univariate and multivariate research that should be included in a published study. The lists may appear long in some cases, but in reality, most of the information can be contained in a few short sentences. At any rate, it is better to include too much information than too little.

General Guidelines for All Research Procedures

Some reporting procedures are unique to univariate or multivariate research, but there are several basic types of information that are relevant in either case. The following are data that should be included in *all* research reports.

1. *Rationale for the research.* The rationale describes why the research was conducted. It usually includes a short statement explaining how the results may add to the field of knowledge. These statements are naturally more detailed for studies in uncharted areas.

2. *Variables used in the analysis.* This includes a description of both independent and dependent variables, explaining how the variables were selected for the study, what marker variables (see Chapter 2) were included, and how extraneous variables were controlled. Each variable also requires some form of justification for its use—variables cannot be added without reason. In addition, the mean and standard deviation for each variable should be reported when necessary.

3. *Sample size.* The researcher should not only state the number of subjects or units of study, but also explain how these entities were selected. Additionally, any departure from normal randomization must be described in detail.

4. *Sample characteristics.* The sample should also be described in terms of its demographic, lifestyle, or other descriptor characteristics. When human subjects are used, at least their age and sex should be indicated.

5. *Methodology.* Every research report requires a description of the methods used to collect and analyze data. The amount of methodological description that should be included depends on the audience; articles written for journals, for instance, must contain more detailed information than reports prepared in private sector research.

6. *Data manipulation.* Often the collected data are not normally distributed, and researchers must use data transformation to achieve an approximation of normality. If such a procedure is used, a full explanation should be given.

7. *Results.* A brief, but thorough, explanation of the results is required in every report. All tables, graphs, charts, and other data displays must be presented in their most parsimonious form.

8. *Summary and conclusions.* In this final section, the researcher summarizes the results and provides an interpretation of the analysis. He or she may also suggest new directions for research in the area.

Additionally, a research report should, where appropriate, include such information as the level of significance, the results of any power analysis, and the type of posttest conducted. All of these types of data are necessary to ensure that readers will be able to understand a study from the description provided in the report.

One final point that should be stressed is that reports of any type of research (except perhaps for privately funded research or studies done for business purposes) should include a statement of some sort concerning where the raw data or other relevant materials may be obtained. The easiest way to store data from a study is to keep a copy of the computer printout of the analysis. Most computer programs include an option that allows the raw data to be printed; using this option, an entire study can be condensed into only a few pages.

Reporting Multivariate Research Studies

Reports of research involving multivariate statistical procedures tend to be more detailed than those of univariate studies since a variety of subjective judgments are involved. This subsection describes the information that should be included in reports dealing with any of the four multivariate methods discussed in Chapter 12. The suggestions provided herein are meant to serve as a basic guide from which researchers can work; additional data relevant to a particular study can be added when necessary.

Factor analysis. Factor analysis is one of the most misused and misreported procedures in mass media research. This is because researchers often do not know enough about factor analysis before using it in research and do not provide enough details regarding its use in their reports. Smith and Blashfield (1978) performed an analysis of reports involving factor analysis that appeared in five broadcast journals between 1964 and 1975. The authors concluded that "factor analytical applications have been imprecisely reported ... [which] can only have a negative effect on theory building in media research and on attempts to further understand this complex set of procedures."

In addition to the information discussed in the previous subsection, researchers reporting factor analyses should state or describe:

1. The type of matrix that was factored (such as correlation or covariance) and where it may be obtained for replication
2. The model used for factor extraction, such as principal components or principal factors
3. The psychometric adequacy of the data, as measured by Bartlett's Sphericity Test, Kaiser's Measure of Sampling Adequacy, or some other test for matrix quality (the two tests mentioned are easy to compute by hand)
4. The method used to estimate the initial communalities, if the principal factors model was used
5. The type of rotational procedure (such as orthogonal or oblique) used to make the data more meaningful for interpretation
6. The criterion used to determine the number of factors to rotate
7. The formula used to compute factor scores (if such scores are used in the study)
8. The cut-off values for salient variables, the elements comprised by a complex factor loading, and the number of salient variables required to constitute a significant factor (usually three)
9. The procedures used to compare the factor structures of different samples or studies, if this is part of the research methodology
10. The factors that were extracted and the relevance of significant variables to each of them
11. The computer program used to perform the factor analysis (since programs differ as to how the results are computed)

Most of this information will be applicable to every factor analysis report. If the study is a replication, a careful comparison should be made between the procedures used in each analysis. It may be that results different from those obtained previous studies may be produced solely because different procedures were used.

For examples of factor analysis reports, the reader is referred to Shinar, Tomer, and Biber (1980) and Alwitt and others (1980).

Canonical correlation. The ease with which the results of a canonical correlation can be reported is directly related to the computer program used for the analysis. Some programs, such as those used by Finn (1979), Cooley and Lohnes (1971), and Veldman (1967), provide all the necessary information in a concise, easy-to-read form. Other programs, such as SPSS and BMD, are less than satisfactory in this regard (although there is nothing inherently wrong with the computational formulas and procedures they employ).

Reporting a canonical correlation is not as complicated as reporting a factor analysis, since fewer procedural and subjective decisions are involved. The following data should be presented in a canonical correlation report:

1. The values for the canonical correlation (R_c) and the canonical correlation squared (R_c^2) (usually included in a table displaying the results of the canonical analysis)
2. All significant canonical component loadings and their associated beta weights (for journal articles, it is best to submit the entire canonical structure —all the roots in the analysis—and allow the editor to decide what information to include)
3. The chi-square tests for significant (if not all) canonical roots and the appropriate levels of significance and degrees of freedom
4. The cut-off values for the canonical correlations, the canonical components and the redundancy indexes

Additionally, since canonical correlation is used sparingly in mass media research, it may be desirable to provide a brief summary describing the purpose of this procedure and how interpretation is accomplished.

Examples of canonical correlation reports can be found in Eastman (1979) and Wimmer (1976).

Discriminant analysis. Information that should be presented in a report of a discriminant analysis includes:

1. Eigenvalues and the canonical structure associated with each discriminant function
2. Wilk's Lambda indicating the discriminatory power of the variable sets and the associated chi-square tests of significance (showing the number of independent functions along which the group centroids or means differ)
3. A visual plot of the centroids (if only one function is significant, researchers should plot the second function as well to allow better visual representation)
4. The variable labels and weights of the discriminant function coefficients
5. Discriminatory power, or amount of variability of the discriminant function that is due to group differences (see Tatsuoka 1970), if required

For examples of discriminant analysis, the reader should consult Tankard and Harris (1980), and Lull, Johnson, and Sweeney (1978).

Multivariate analysis of variance. MANOVA is the least controversial of the four multivariate methods described in this book, basically because it entails fewer subjective judgments with regard to procedural steps, cutoff levels, and the interpretation of results. Most discussions of MANOVA in multivariate textbooks are brief because the statistical principles involved are well established and hence do not require elaborate justifications. In addition to the basic information included in all research reports, a MANOVA report should include:

1. The correlation between the dependent measures
2. Alpha levels used to determine significance
3. Means and variance for each dependent measure
4. Univariate and multivariate results associated with each hypothesis in the study
5. Results of any posttest analysis

A good example of a MANOVA report is provided by Talley and Richmond (1980).

The Mechanics of Writing a Research Report

Given the wide variety of approaches to research, it stands to reason that the possible approaches to writing a research report are equally varied. However, some comfort may be derived from the fact that most research reports include five basic sections or chapters: (1) an introduction; (2) a review of literature and/or a justification; (3) a description of methodologies; (4) a presentation of results; and (5) a summary or conclusion. This format is used both in articles for professional publication and in management summaries.

The purpose of this section is to describe some of the basics of writing research reports. Beginning researchers may find the writing style awkward or unaesthetic, but there is a definite purpose behind the rules governing scientific writing: clarity. Every effort must be made to avoid ambiguity.

Although journals' and businesses' writing requirements may vary in some details, the following suggestions are applicable to most types of reports:

1. Avoid using first personal pronouns: I, me, my, we, and so on. Research reports are almost always written in third person. For example, instead of writing, "I selected the subjects randomly . . . ," one should write, "The researcher selected subjects randomly. . . ." First personal pronouns should only be used when the article is a commentary.

2. Place all tables, graphs, charts, and figures on separate pages when submitting a paper for professional publication. This is done not

because it facilitates editing and review, but rather because if the article is accepted, these pages will be typeset by one department of the printing company and the text by another. The two sections are then combined shortly before the article goes to print. In management reports, tables, graphs, and other displays are included in the text unless they are too large, in which case they should be placed on separate pages.

3. Always read the guidelines published by each journal for submitting articles. They provide specific rules concerning acceptable writing style, footnote and bibliography formats, the number of copies to submit, and so forth. A researcher who fails to follow these guidelines may decrease the chance that his or her report will be accepted for publication—or at least substantially delay the process while he or she corrects the errors.

4. Be stylistically consistent with regard to tables, charts, graphs, section headings, and so forth. Tables, for example, should follow the same format and be numbered consecutively.

5. Clearly label all displays with meaningful titles. Each table, graph, chart, or figure caption should accurately describe the material presented and its contribution to the remainder of the report.

6. Keep language and descriptions as simple as possible by avoiding unnecessary and overly complex words, phrases, and terms. The goal of scientific writing is to explain findings clearly, simply, and accurately.

7. An executive summary is required for management reports. This is essentially an abstract—a brief report of the study's highlights. The summary is usually the first piece of information presented in the report, to give the reader an idea of what type of information is contained in the remainder of the report.

8. Use the active rather than passive voice. For example, instead of writing, "It was found by the researchers that . . . ," one should write, "The researchers found that. . . ." Writing in the active voice makes reading more pleasant and also requires fewer words.

9. Proofread the manuscript carefully. Even researchers who are meticulous in their scientific approach can make errors in compiling a manuscript. All manuscripts, whether intended for publication or for management review, should be proofread several times to check for accuracy.

10. Miscellaneous considerations:

   a. Use "fair" language; that is, avoid phrases or references that could be interpreted as sexist or racist. Editors should be consulted with regard to the use of masculine pronouns.
   b. Check all data for accuracy. Even one misplaced digit may affect the results of a study.
   c. Use acceptable grammar.

d. Provide acknowledgments whenever another researcher's work is included in the report.
e. Include footnotes to indicate where further information can be obtained.

Research Ethics

The majority of mass communication research involves observations of human beings—asking them questions or examining what they have done. Since human beings have certain rights, it is necessary for the researcher to ensure that the rights of the participants in a project are not violated. This requires a consideration of the ethics involved in particular research situations.

Ethics are concerned with distinguishing right from wrong and the proper from the improper. Unfortunately, there are no universal definitions for these terms. Instead, a series of guidelines, broad generalizations, and suggestions have been endorsed or at least tacitly accepted by those in the research profession. While these guidelines may not provide an answer to every ethical question that may arise, they at least help make researchers more sensitive to the issues involved.

Before discussing these specific guidelines, this section will examine some hypothetical research situations that might involve ethical problems:

1. A researcher at a large university is interested in describing the media usage pattern of students. The researcher hands out questionnaires to the students in his or her introductory mass media course and tells the students that if they do not fill out the questionnaire, they will lose points toward their grade in the course.

2. A researcher is conducting a mail survey about attendance at X-rated motion pictures. The questionnaire states that responses will be anonymous. Unknown to the respondent, each return envelope is marked with a code that enables the researcher to identify the sender.

3. A researcher recruits subjects for an experiment by stating that the experiment involves watching "a few scenes from some current movies." Those who decide to participate are shown several scenes of bloody and graphic violence.

4. A researcher shows one group of children a violent television show and another group a nonviolent program. After viewing, the children are sent to a public playground, where they are told to play with the children who are already there. The researcher records each instance of violent activity performed by the experimental and control children.

5. Subjects in an experiment are told to submit a sample of their newswriting to an executive of a large newspaper. They are led to believe that the subject who submits the best work will be offered

a job at the paper. In fact, the so-called "executive" is a confederate in the experiment and severely criticizes everyone's work.

These examples should be kept in mind while reading the following guidelines to ethics in mass media research.

Voluntary Participation and Informed Consent

An individual is entitled to decline to participate in any research project or to terminate participation at any time. Participation in an experiment, survey, or focus group is always voluntary, and any form of coercion is unacceptable. Researchers who are in a position of authority over subjects (as in Situation 1 above) should be especially sensitive to *implied* coercion: even though the researcher might tell the class that participation is voluntary and will not affect their grades, many students may still believe that it will. In such situations, it is advisable to keep the questionnaires anonymous and to have the person in authority be absent from the room while the survey is administered.

Voluntary participation is less of an ethical problem in mail and telephone surveys, since respondents are free to hang up the phone or to throw away the questionnaire. Nonetheless, a researcher should not attempt to induce subjects to participate by misrepresenting the organization sponsoring the research or by exaggerating its purpose or importance. For example, phone interviewers should not be instructed to identify themselves as representatives of the "Department of Information" so as to mislead people into thinking the survey is government-sponsored. Likewise, mail questionnaires should not be constructed to mimic census forms, tax returns, social security questionnaires, or other official government forms.

Closely related to voluntary participation is the notion of *informed consent*. In order for people to volunteer for a research project, they need to know enough about the project to make an intelligent choice. Researchers have the responsibility to inform potential subjects or respondents of all features of the project that can reasonably be expected to influence participation. Respondents should understand that an interview may take as long as 45 minutes, or that a second interview is required, or that upon completing a mail questionnaire, they may be singled out for a telephone interview.

In an experiment, informed consent means that potential subjects must be warned of any possible discomfort or unpleasantness that might be involved. Subjects should be told if they are to receive or administer electric shocks, be subjected to unpleasant audio or visual stimuli, or undergo any procedure that may cause concern. Informed consent also includes describing any unusual measurement techniques that may be used. Researchers have an obligation to answer candidly and truthfully, as far as possible, all of the participant's questions about the research.

Experiments that involve deception (see the following subsection) cause special problems with regard to obtaining informed consent. If deception is absolutely necessary to conduct an experiment, is the experimenter then obligated to inform the subject that he or she may be deceived during the upcoming experiment? Will such a disclosure affect participation in the experiment? Will it also affect the experimental results? Should one compromise by telling potential subjects that deception will be involved for some of them but not for others?

A second problem is deciding exactly how much information about a research project must be disclosed in order to achieve informed consent. Is it enough to explain that the experiment involves rating commercials, or is it necessary to add that the experiment is designed to test whether subjects with high IQs prefer different commercials from those with low IQs? Obviously, there are situations where the researcher cannot reveal everything about the project for fear of contaminating the results. For example, if the goal of the research is to examine the influence of peer pressure on commercial evaluations, alerting the subjects to this fact might change their behavior in the experiment.

Finally, one must consider the form of consent to be obtained. Many organizations, such as schools, require that written consent be secured from the system superintendent, the school's principal, and the parents of any children who participate in the research. Written consent is also becoming a requirement for many government-sponsored research programs as well as by increasing numbers of university research committees. But what constitutes consent to a personal interview or a telephone survey? Does the fact that the respondent provides answers to the interviewer imply consent? Or is it necessary to have the respondent sign a written consent form before the questioning actually begins?

Concealment and Deception

Concealment and deception techniques are encountered most frequently in experimental research. Concealment involves withholding certain information from the subjects; deception involves deliberately providing them with false information. Both raise certain ethical problems. The difficulty in obtaining consent has already been mentioned. A second problem derives from a general feeling that it is inherently wrong for experimenters to lie or otherwise deceive subjects. Many critics argue that deception transforms a subject from a human being into a manipulated object, and is therefore demeaning to the participant involved. Moreover, once subjects have been deceived, they are likely to expect to be deceived again in other research projects. It is claimed that deception creates an atmosphere of mistrust between subject and researcher and that subjects may actually attempt to turn the tables by trying to deceive the experimenter.

On the other hand, researchers who favor deception argue that in some instances deception is the only possible way to conduct a study. Additionally, they claim that the harm done to those who are deceived is outweighed by the benefits of the research to scientific knowledge. The same arguments can also be used both for and against concealment. In general, however, concealment is somewhat less of an ethical problem, provided that enough information is given to subjects to allow informed consent and that all of the subjects' questions are answered candidly.

Obviously, deception is not a technique that should be used indiscriminately. Kelman (1967) suggested that before the investigator settles on deception as an experimental tactic, three questions should be examined:

1. How significant is the proposed study?
2. Are alternative procedures available that would provide the same information?
3. How severe is the deception? (It is one thing to tell a subject that the experimentally constructed message he or she is reading was taken from the *New York Times*; it is another to report to a subject that the test he or she has just completed was designed to measure latent suicidal tendencies.)

When an experiment is concluded, especially one involving concealment or deception, it is the responsibility of the investigator to "debrief" subjects with regard to the experiment in which they have participated. Debriefing should be thorough enough to remove any lasting effects that might have been created by the experimental manipulation or by any other aspect of the experiment. Subjects' questions should be answered and the potential value of the experiment stressed.

Protection of Privacy

The problem of protecting the privacy of participants usually occurs more often in survey research than in laboratory studies. Subjects have a right to know whether their privacy will be maintained and who will have access to the information they provide. There are two ways to guarantee privacy: (1) by assuring anonymity; and (2) by assuring confidentiality. A promise of anonymity is a guarantee that a given respondent cannot possibly be linked to any particular response. In many research projects anonymity is actually an advantage, since it encourages respondents to be honest and candid in their answers. Strictly speaking, personal and telephone interviews cannot be anonymous, since the interviewer can link a given questionnaire to a specific person, household, or telephone number. In such instances, the researcher should provide a promise of confidentiality; that is, even though the respondents can be identified, the information they provide will never be made public. A researcher should never use the term *anonymous* to mean "confidential."

Additionally, respondents should be told who *will* have access to the information they provide. The researcher's responsibility for assuring confidentiality does not end once the data have been analyzed and the study is concluded. Questionnaires that identify persons by name should not be stored in public places, nor should other investigators be given permission to examine confidential data unless all identifying marks have first been obliterated.

"Do No Harm"

As a general ethical guideline, the admonition to "do no harm," borrowed from the medical field, is appropriate. Researchers should never injure any of the people they study, nor should they create circumstances where innocent bystanders could be harmed. Displaying a violent film to a group of children and then turning them loose on a group of unsuspecting playmates raises obvious ethical problems. Researchers must also avoid subjecting participants to more subtle forms of psychological harm. Any situation where respondents are asked to reveal potentially embarrassing personal information should be handled with extreme care (if it cannot be avoided altogether).

Ethics in Data Analysis and Reporting

Researchers are also responsible for maintaining professional standards in the way they analyze and report their data. The ethical guidelines in this area are less controversial and more clear-cut. One cardinal rule is that researchers have a moral and ethical obligation to refrain from tampering with data: questionnaire responses and experimental observations may not be fabricated, altered, or discarded. Similarly, researchers are expected to maintain reasonable care in processing the data in order to guard against needless errors that might affect the results.

Researchers should never conceal any information that might influence the interpretation of their findings. For example, if a two-week delay occurred between the testing of the experimental group and the testing of the control group, this fact should be reported so that other researchers might be aware of the effects of history and maturation on the results. A full and complete description of method, and particularly of any departure from standard procedures, should be provided in every research report.

Since science is a public activity, researchers have an ethical obligation to share their findings and methods with other researchers. All questionnaires, experimental materials, measurement instruments, instructions to subjects, and other relevant items should be made available to those who wish to examine them.

Finally, all investigators are under an ethical obligation to draw conclusions from their data that are consistent with those data. Interpretations should not be stretched or distorted to fit a personal point of view or a favorite theory, or to gain or maintain a client's favor. Nor should researchers attribute greater significance or credibility to their data than they justify. For example, when analyzing correlation coefficients obtained from a large sample, it is possible to achieve statistical significance with an r of only, for example, .10. While it is thus perfectly acceptable for the investigator to report a statistically significant result, he or she should also mention that the predictive utility of the correlation is not large, as it explains only 1% of the total variation involved. In short, researchers should report results with candor and honesty.

A Researcher's Code of Ethics

Formalized codes of ethics have yet to be developed by all professional associations involved in mass communication research. One research organization that has developed its own ethical code is the American Association for Public Opinion Research. The code is reproduced below.

American Association for Public Opinion Research

Code of Professional Ethics and Practices

I. Principles of Professional Practice in the Conduct of Our Work
   - A. We shall exercise due care in gathering and processing data, taking all reasonable steps to assure the accuracy of results.
   - B. We shall exercise due care in the development of research designs and in the analysis of data.
     1. We shall recommend and employ only research tools and methods of analysis which, in our professional judgment, are well suited to the research problem at hand.
     2. We shall not select research tools and methods of analysis because of their capacity to yield a misleading conclusion.
     3. We shall not knowingly make interpretations of research results, nor shall we tacitly permit interpretations, which are inconsistent with the data available.
     4. We shall not knowingly imply that interpretations should be accorded greater confidence than the data actually warrant.
   - C. We shall describe our findings and methods accurately and in appropriate detail in all research reports.

II. Principles of Professional Responsibility in Our Dealings with People
   - A. The Public:
     1. We shall cooperate with legally authorized representatives of the public by describing the methods used in our studies.

2. When we become aware of the appearance in public of serious distortions of our research we shall publicly disclose what is required to correct the distortions.

B. Clients and Sponsors:

1. When undertaking work for a private client we shall hold confidential all proprietary information obtained about the client's business affairs and about the findings of research conducted for the client, except when the dissemination of the information is expressly authorized by the client or becomes necessary under terms of Section II-A-2.

2. We shall be mindful of the limitations of our techniques and facilities and shall accept only those research assignments which can be accomplished within these limitations.

C. The Profession:

1. We shall not cite our membership in the Association as evidence of professional competence, since the Association does not so certify any persons or organizations.

2. We recognize our responsibility to contribute to the science of public opinion research and to disseminate as freely as possible the ideas and findings which emerge from our research.

D. The Respondent:

1. We shall not lie to survey respondents or use practices and methods which abuse, coerce, or humiliate them.

2. Unless the respondent waives confidentiality for specified uses we shall hold as privileged and confidential all information that tends to identify a respondent with his or her responses. We shall also not disclose the names of respondents for non-research purposes.

Finding Support for Mass Media Research

Research costs money. Finding a source for research funds is a problem that confronts both quantitative and qualitative researchers in all fields of mass communication. This section contains a listing of organizations that have supported mass communication research projects. A researcher in need of funding should contact these organizations for specific details as to the types of studies they support and the amount of funds available, as well as instructions for preparing research proposals.

Researchers at a university or college should determine whether their institution has a program of research grants for individual faculty members. Many colleges award such grants, often on a competitive basis, for research in mass communication. Typically these grants are modest in size—usually under $5,000—but they are among the easiest to apply for and to administer. In many cases grants are also available for student research.

There are several philanthropic foundations that sponsor mass media research. Among the better known are the Ford Foundation, the John and Mary Markle Foundation, and the Alfred P. Sloan Foundation. The amounts these organizations give to support research vary greatly, but they generally range from about $5,000 to as much as $150,000. Competition is stiff, and the researcher should be certain that his or her research area is one for which these foundations provide funding. As an alternative, there may be smaller foundations located near the researcher's base of operations that could be investigated.

Certain departments of the U.S. government sponsor mass communication research (although as of 1982, many were expecting budget cuts). Among the departments that have been active in supporting media research are the National Institute of Mental Health, the National Science Foundation, and the National Endowment for the Humanities. Other possible funding agencies can be identified by looking through the *Federal Grant Register.* Applying for a government grant tends to be complicated, and there are many guidelines and regulations that the researcher must follow. In addition, there is the usual problem of government red tape. Nonetheless, these agencies have been known to make sizable grants to investigators.

Many professional media associations sponsor continuing programs to support research relevant to their particular field. In radio and television, the National Association of Broadcasters awards annual grants for research in broadcasting. The competition is keen: approximately half a dozen grants of $1,200 are made to individuals each year. This program is open to both professors and students interested in broadcasting research. In the print media, the American Newspaper Publishers Association (ANPA) sponsors research having to do with readership and circulation. The ANPA has a News Research Center located at Syracuse University. The Frank E. Gannett Newspaper Foundation also provides research funding, and the Magazine Publishers Association (MPA) sponsors magazine-related research.

The American Academy of Advertising sponsors an annual competition to award a $1,000 grant to a new faculty member (one who has been teaching less than five years) to conduct research studies in the field of advertising. The American Association of Advertising Agencies has also funded research projects. Similarly, the Public Relations Foundation has a program of small grants ($1,000–$1,500) available to support research in public relations.

Many researchers have obtained money to finance their research by working with the media industry. The three television networks have research departments that are willing to examine proposals from outside investigators that might be of interest to them. Occasionally, they will even sponsor a research program themselves. The American Broadcasting Company in the mid-1970s funded five separate research projects submitted by academic researchers, providing $20,000 for each. The Columbia Broadcasting System recently sponsored a lengthy audience survey conducted among British youngsters. In addition, the larger group owners in broadcasting also have research departments that might be approached. The Corporation for Public Broadcasting, for example, has sponsored several audience studies relating to their programs. Similarly, large newspaper chains are potential funding sources. In the public relations field, many researchers have obtained support by contacting the professional organization of the industry they are studying or by working with a private company or corporation.

Industry support can be a mixed blessing. On the one hand, working with industry backing makes it easier for a researcher to enlist the cooperation of people or organizations within that industry and also facilitates obtaining data about the inner workings of the industry. On the other hand, many media industries are interested in limited research areas that may not have much theoretical attraction for the researcher. A company may insist that it specify the focus of the study and the variables to be examined. Therefore, when approaching a media organization or any other private company for support, it is wise to determine in advance what control, if any, the private organization will have over the design, execution, and subsequent publicizing of the project.

Summary

Writing a research report is naturally an important step in the scientific process, since the report places the research study in the public domain for consideration and confirmation. Beginning researchers generally find the process much easier after they have completed one or two studies. A key to successful writing is to follow the guidelines developed by journal editors, or styles developed by individual companies or businesses. The same basic five-part format is used for all reports.

Ethical considerations in conducting research should not be overlooked by investigators. Nearly every research study has the potential of affecting subjects in some way, either psychologically or physically. Researchers dealing with human subjects must take great care to ensure that all precautions are taken to alleviate any potential harm to subjects. This includes carefully planning a study as well as debriefing subjects upon completion of a project.

The final part of Chapter 17 describes financing research projects. This topic is relevant to all researchers since cost often cancels many good research topics. The chapter describes a variety of sources which provide financial assistance; none should be overlooked for help.

Questions and Problems for Further Investigation

1. Read a recent article from an academic journal that publishes mass media research. See how well the authors follow the reporting guidelines discussed in this chapter.
2. Using the five research examples described at the beginning of the section dealing with ethics, provide, if possible, an alternate way of conducting the study that would be ethically acceptable.

References and Suggested Readings

Alwitt, L. F., Anderson, D. R., Lorch, E. P., and Levin S. R. 1980. "Preschool Children's Visual Attention to Attributes of Television." *Human Communication Research* 7:52–67.

Cooley, W. W., and Lohnes, P. R. 1971. *Multivariate Data Analysis.* New York: Wiley.

Dixon, W. J., ed. 1978. *Biomedical Computer Programs.* Berkeley: University of California Press.

Eastman, S. T. 1979. "Uses of Television Viewing and Consumer Lifestyles: A Multivariate Analysis." *Journal of Broadcasting* 23:491–500.

Finn, J. 1979. *Multivariance.* Chicago: National Educational Resources.

Kelman, H. 1967. "Human Use of Human Subjects: The Problem of Deception in Social Psychological Experiments." *Psychological Bulletin* 67: 1–11.

Lull, J. T., Johnson, L. M., and Sweeney, C. E. 1978. "Audiences for Contemporary Radio Programs." *Journal of Broadcasting* 22:439–453.

Nie, N., *et al.* 1975. *Statistical Package for the Social Sciences.* New York: McGraw-Hill.

Rummel, R. J. 1970. *Applied Factor Analysis.* Evanston: Northwestern University Press.

Shinar, D., Tomer, A., and Biber, A. 1980. "Images of Old Age in Television Drama." *Journal of Communication* 30:50–55.

Smith, J. R., and Blashfield, R. K. 1978. "Reporting Factor Analyses in Mass Media Research: A Review of Methods." *Journal of Broadcasting* 22: 187–98.

Talley, M. A., and Richmond, V. P. 1980. "The Relationship between Psychological Gender Orientation and Communication Style." *Human Communication Research* 6:326–39.

Tankard, J. W., and Harris, M. C. 1980. "A Discriminant Analysis of Television Viewers and Non-Viewers." *Journal of Broadcasting* 24:399–409.

Tatsuoka, M. M. 1970. *Multivariate Analysis.* New York: Wiley.

Veldman, D. 1967. *Fortran Programming for the Behavioral Sciences.* New York: Holt, Rinehart and Winston.

Wimmer, R. D. 1976. "A Multivariate Analysis of the Uses and Effects of the Mass Media in the 1968 Presidential Campaign." Unpublished doctoral dissertation, Bowling Green State University.

Wimmer, R. D., and Reid, L. N. 1982. "Willingness of Communication Researchers to Respond to Replication Requests." *Journalism Quarterly* (in press).

APPENDIX 1

Tables

Table 1. Random numbers

0	8	9	5	6	4	4	8	9	4	0	7	5	9	7	0	4	5	3	1	2	7	8	6	6
8	2	4	4	8	8	0	2	6	5	5	0	3	5	9	1	3	8	6	8	8	3	1	8	5
3	1	2	3	7	6	4	1	1	4	3	5	2	7	4	9	3	2	7	5	5	4	7	6	2
2	3	8	1	8	6	6	1	0	8	4	1	0	5	0	4	8	5	3	7	8	7	6	5	7
0	0	4	3	6	5	5	2	3	5	2	4	3	3	9	3	2	5	2	0	8	4	6	2	1
1	2	8	9	7	5	8	9	7	8	6	7	4	0	4	0	4	9	7	8	5	0	2	9	8
9	8	4	6	9	9	0	8	0	2	3	2	8	0	5	4	5	0	6	7	6	2	3	9	8
0	7	3	6	9	5	1	6	3	8	0	5	9	0	0	2	0	9	3	6	8	8	2	4	3
2	2	3	9	5	7	9	4	0	6	7	3	6	9	6	4	1	7	3	6	5	1	8	2	6
4	9	5	6	9	3	1	4	7	8	1	5	6	7	2	2	4	6	3	6	5	4	2	1	2
4	0	6	6	8	5	4	3	7	8	3	2	6	8	1	2	2	7	0	6	5	3	5	8	4
6	3	3	2	0	3	9	7	0	2	3	6	9	5	3	4	1	6	1	8	3	9	4	3	3
0	6	1	8	4	2	1	8	6	7	5	4	1	9	0	3	2	4	1	5	7	7	4	0	8
2	2	4	2	9	6	8	5	8	2	6	1	0	7	6	1	7	9	2	0	9	2	8	7	8
8	3	2	3	0	7	4	3	5	8	9	0	8	0	5	8	8	7	1	3	6	0	1	3	9
2	3	1	8	2	3	1	0	9	0	0	8	9	1	2	0	3	7	0	2	0	1	8	1	7
0	8	7	3	4	4	5	1	8	7	4	5	1	9	9	0	3	2	2	3	1	2	6	4	6
5	8	5	6	7	6	1	0	1	6	7	0	2	1	9	1	6	3	2	0	1	1	5	5	9
6	1	1	0	5	1	3	6	7	7	7	8	2	4	5	9	3	0	7	6	7	9	1	1	6
5	3	6	1	2	7	2	6	2	7	3	3	6	8	2	6	5	5	8	4	2	4	2	1	8
8	7	3	9	5	1	1	8	4	1	8	5	6	6	0	6	9	2	2	6	8	2	5	8	5
2	9	1	9	9	5	6	1	8	6	6	4	0	5	0	0	8	8	2	5	9	2	0	1	2
8	1	0	2	1	7	2	0	2	7	6	8	4	8	0	2	6	2	8	0	8	3	6	0	7
9	7	1	5	5	7	4	6	1	5	6	5	9	9	2	2	7	1	2	7	0	0	5	0	9
6	3	7	9	8	8	7	4	9	5	0	3	3	0	3	7	0	7	5	8	1	2	8	3	1
9	4	2	2	1	3	2	0	5	6	0	6	0	9	0	9	3	1	7	8	1	2	3	1	1
5	2	8	5	1	0	2	4	6	0	8	3	4	2	9	0	2	4	0	5	2	7	8	8	8
7	9	7	1	3	7	2	4	6	3	8	4	0	2	5	5	4	6	1	6	5	4	6	3	0
0	1	5	0	6	5	1	1	8	0	9	4	1	1	2	6	1	4	2	0	8	6	3	1	0
5	8	1	7	4	7	5	6	2	1	9	3	7	4	0	4	6	4	6	9	6	7	5	0	6
2	5	0	7	5	1	6	0	4	0	4	1	9	4	9	8	3	6	3	8	0	0	1	7	9
8	8	3	7	8	1	4	6	3	8	0	5	6	4	4	3	5	0	6	9	5	5	0	6	0
4	3	1	8	7	3	4	1	7	1	6	1	5	2	7	9	4	0	2	9	9	6	8	7	6
9	1	4	7	7	4	3	7	4	2	5	5	0	2	1	1	1	4	0	6	4	7	5	9	6
8	6	0	8	2	9	3	4	3	4	7	6	9	6	1	8	2	3	3	8	3	4	6	8	3
3	3	0	6	2	3	8	7	4	3	8	3	1	1	5	9	7	4	4	4	9	7	6	0	9
1	8	2	0	2	9	8	8	0	1	6	8	0	7	5	6	0	8	3	9	2	1	1	2	0
4	7	4	1	1	8	5	9	6	9	7	7	8	0	8	0	8	5	7	2	6	9	4	6	7
7	2	8	1	1	0	4	0	5	0	0	8	2	5	7	4	9	4	0	6	9	7	1	8	0
8	4	0	0	8	1	8	7	1	5	0	1	3	7	3	1	1	4	1	9	7	1	7	8	5
1	5	0	5	3	1	9	7	5	0	3	7	6	3	4	7	2	2	0	5	0	0	7	5	1
6	8	5	1	2	4	1	0	4	6	2	5	9	9	3	2	5	6	0	1	2	0	6	7	7
7	6	5	5	4	6	1	9	1	1	7	9	9	9	6	6	7	1	3	7	7	4	8	8	2
7	8	2	4	2	1	6	4	3	9	7	2	6	6	5	7	0	1	2	8	9	7	1	4	5
9	0	3	3	8	1	3	5	1	4	2	8	7	7	0	3	5	8	0	8	4	2	6	6	4
5	5	4	8	6	5	6	8	0	3	2	0	4	8	4	5	6	6	5	4	7	1	3	1	2
0	6	4	9	7	7	9	8	0	6	4	0	9	2	4	7	8	2	5	1	7	2	3	5	2
6	0	6	7	8	0	8	7	6	8	5	0	1	3	4	3	0	4	7	0	5	2	4	1	3
1	6	3	6	4	9	6	5	3	5	5	3	0	3	3	8	3	7	9	1	1	5	8	2	2
2	1	5	9	7	1	2	6	4	4	5	0	2	1	4	5	1	1	7	0	4	0	1	3	0

Table 1. Random numbers (cont.)

5	0	3	9	1	8	3	8	9	5	5	6	7	3	0	6	7	9	7	1	4	9	2	3	3
3	5	8	1	8	1	6	3	4	7	0	6	7	7	8	9	6	2	0	8	5	0	4	3	7
7	0	6	4	0	6	9	0	5	9	3	3	7	7	1	1	4	4	3	8	0	6	2	1	8
1	0	4	9	2	7	8	1	6	4	4	9	3	2	9	6	7	3	2	4	2	6	4	9	6
7	7	7	0	3	2	5	7	9	3	0	5	6	6	5	8	7	6	2	8	5	2	5	3	8
3	1	4	2	0	1	2	3	5	8	0	4	9	9	9	5	6	4	8	6	4	3	5	0	8
8	7	9	8	4	6	4	1	7	0	8	6	0	0	6	1	7	0	9	0	2	9	8	4	2
5	0	6	9	7	6	4	6	4	9	6	6	0	5	3	2	7	9	2	4	4	4	0	6	5
0	9	7	6	2	3	7	3	6	5	7	7	4	8	5	9	4	9	6	6	0	9	5	6	3
1	1	2	9	9	4	6	0	0	6	3	7	1	3	1	9	1	2	6	6	0	8	7	5	2
9	5	5	5	1	9	7	5	9	0	3	2	1	5	6	1	1	1	2	8	3	5	9	5	5
5	6	2	2	6	5	2	0	4	0	5	8	1	8	6	1	2	3	9	0	3	4	3	0	3
3	0	8	5	5	8	7	5	1	7	1	0	7	0	2	7	4	9	9	5	4	9	3	4	6
1	9	4	1	2	5	8	1	2	4	4	9	7	5	9	7	5	8	8	6	2	2	2	4	0
1	6	0	1	7	5	6	9	4	1	7	3	2	2	6	5	1	4	5	9	8	9	9	2	4
9	4	3	4	6	5	3	2	3	0	8	5	6	6	1	1	0	6	6	6	9	6	0	1	1
3	8	5	2	2	5	3	1	3	4	8	8	2	8	7	5	4	6	4	6	4	0	3	3	4
6	5	9	8	7	5	1	5	0	1	3	1	3	5	7	1	1	7	6	6	6	6	8	4	5
9	9	7	6	9	8	8	7	0	6	1	5	7	9	7	1	5	9	7	9	2	6	7	1	1
3	2	8	0	3	7	7	6	8	3	1	2	6	3	0	8	1	4	8	6	1	2	6	6	8
8	9	9	2	9	7	7	4	2	3	3	5	9	2	3	5	8	6	7	3	0	6	4	9	9
5	2	2	0	3	2	8	7	3	4	1	2	6	8	9	6	8	9	4	1	7	6	8	2	9
9	3	7	1	9	8	3	6	0	2	8	6	3	5	3	0	1	6	1	3	3	8	3	4	8
0	6	7	9	9	0	3	7	7	2	6	0	7	7	1	1	8	1	2	9	9	7	8	0	6
6	5	3	1	0	4	2	4	5	1	4	9	5	3	9	0	2	2	4	5	9	9	9	0	0
4	1	8	9	1	7	4	3	6	4	4	6	6	6	0	7	6	3	2	5	8	2	0	6	8
4	5	4	7	1	1	4	5	0	4	7	9	4	0	6	1	2	1	9	4	9	9	0	2	3
2	5	4	3	3	6	3	1	4	0	9	3	7	9	1	1	8	8	1	8	0	3	1	9	5
4	3	6	4	0	1	7	8	2	0	4	9	5	9	7	9	0	3	3	7	2	9	9	4	0
2	3	8	5	4	4	3	3	0	6	1	0	7	3	5	3	1	3	2	0	6	0	9	1	7
1	6	4	8	7	9	9	9	1	3	1	0	8	6	7	5	6	9	0	3	1	6	8	2	0
4	8	1	6	3	4	5	0	2	7	5	7	0	8	3	2	4	8	5	3	2	9	6	8	1
4	2	1	9	4	6	2	3	0	1	1	6	1	0	7	2	2	3	4	8	7	9	1	4	6
4	0	7	6	5	4	2	9	5	3	3	9	0	6	3	0	2	5	4	9	5	3	6	0	8
8	4	9	3	0	8	2	8	4	0	4	5	6	9	0	6	8	1	1	4	6	7	4	8	1
1	7	6	3	8	1	4	6	2	2	9	4	5	0	3	5	7	0	0	2	4	1	7	1	2
5	6	4	6	9	0	1	5	1	5	5	0	3	1	4	5	1	2	7	0	2	4	9	9	6
0	3	6	0	7	1	4	8	0	3	5	4	8	8	0	4	0	6	7	3	3	1	1	7	4
6	7	2	9	0	4	2	9	2	6	4	6	4	6	4	6	9	4	6	2	3	9	4	8	8
0	3	1	4	5	9	5	0	8	2	6	5	0	8	5	8	0	7	5	0	9	5	3	1	5
7	3	0	9	3	6	1	9	3	1	3	9	8	3	9	7	7	6	6	5	3	0	2	6	8
8	6	7	9	6	6	8	3	4	0	5	9	5	1	7	8	0	1	0	8	9	7	1	4	6
4	9	5	8	6	8	0	4	4	4	5	6	7	4	8	1	7	1	4	9	2	9	5	1	9
6	0	3	9	9	5	8	4	4	1	5	4	0	6	8	6	0	2	0	0	1	8	8	8	0
4	1	0	5	3	6	3	5	0	6	4	0	0	1	2	1	8	2	9	5	4	8	7	2	5
5	2	7	9	6	5	7	4	5	1	3	3	8	8	4	4	0	4	1	8	9	1	1	6	5
3	4	6	1	2	1	8	7	4	7	6	3	3	5	0	0	7	9	1	6	4	0	7	4	6
8	2	2	0	8	8	8	7	3	8	3	1	5	8	4	9	5	1	9	1	7	9	7	9	9
4	8	7	0	7	8	9	4	3	0	9	2	3	5	4	7	2	1	4	6	6	8	6	3	2
9	0	4	3	8	0	1	5	7	6	7	1	6	3	0	5	7	3	7	1	0	9	5	6	6

Table 1. Random numbers (cont.)

8	2	8	9	7	9	6	9	7	9	0	8	2	9	8	1	5	6	9	3	2	9	2	3	3
9	4	6	9	2	6	8	4	4	7	8	3	5	1	0	1	3	9	9	2	9	0	4	0	8
5	6	7	4	2	7	4	1	2	7	3	1	5	8	3	1	0	7	3	8	7	5	2	5	1
8	0	9	9	8	3	2	9	7	5	5	8	0	5	2	1	3	4	2	3	8	6	8	3	6
6	7	0	3	7	9	8	8	2	0	9	1	0	6	0	7	2	4	5	1	3	3	5	1	0
8	1	3	0	0	8	3	4	8	8	3	4	8	9	9	2	0	4	3	9	6	7	6	5	7
1	7	6	2	5	8	6	2	6	6	8	0	8	3	9	8	8	7	4	2	1	3	3	3	2
9	9	7	1	7	5	9	1	3	2	4	6	0	5	9	0	7	3	8	2	3	5	4	7	1
0	4	6	4	0	1	7	9	9	3	6	8	1	5	3	7	1	1	9	5	1	0	1	4	8
9	7	8	2	1	2	9	7	2	0	6	4	2	5	2	7	0	8	1	1	9	7	7	7	0
2	4	6	4	6	3	6	7	5	2	0	0	5	4	7	3	3	4	1	0	7	4	4	0	9
8	5	4	5	4	7	7	4	0	0	5	0	6	4	2	8	8	0	8	0	9	9	0	5	8
5	8	6	7	6	6	4	7	0	1	4	9	9	5	7	2	1	4	1	1	9	7	7	3	5
1	3	8	1	4	7	0	7	4	8	8	4	4	0	1	2	5	1	4	8	1	7	7	3	2
4	1	5	9	7	9	5	6	6	7	4	5	6	1	8	8	8	2	8	9	0	0	9	2	5
9	5	4	7	0	6	8	1	2	1	4	0	4	5	8	3	1	6	0	1	9	7	5	6	0
3	7	2	7	4	1	4	8	3	6	4	1	6	1	9	0	4	1	3	2	6	8	9	2	5
9	7	1	8	1	0	8	3	6	0	1	7	5	0	6	3	2	7	9	2	5	6	2	9	9
9	9	9	9	1	9	4	2	6	9	5	8	5	6	8	3	9	8	6	9	9	6	8	2	5
9	3	0	1	8	1	5	8	8	1	1	4	4	6	6	4	1	0	9	6	6	7	5	5	8
7	9	4	6	8	9	0	6	6	9	5	4	3	1	9	5	1	9	5	6	2	8	2	7	4
3	5	5	4	5	2	5	2	2	1	4	8	2	0	9	1	8	4	3	5	0	3	2	6	5
6	7	2	1	9	0	5	4	3	3	9	8	9	0	1	2	6	6	1	3	0	4	5	4	1
4	0	5	3	9	2	6	3	2	2	0	4	2	0	9	1	0	0	8	8	8	0	2	8	1
2	1	5	7	3	7	3	6	2	8	9	3	2	8	7	9	6	7	9	5	1	9	5	5	4
8	2	9	1	7	6	5	0	5	7	4	2	4	7	5	1	4	2	8	4	0	2	0	4	5
0	4	9	2	5	9	9	8	7	4	7	3	2	2	1	7	7	1	9	5	1	4	4	9	4
3	8	6	7	5	6	1	5	3	0	9	0	8	4	0	4	6	7	2	2	6	8	4	3	5
7	1	8	8	3	6	3	7	4	3	6	3	3	0	1	3	4	9	7	3	8	9	2	3	6
2	3	0	4	7	4	6	9	9	9	8	7	4	4	2	8	1	4	4	4	0	0	6	0	8
8	6	4	4	0	7	1	2	9	6	3	1	3	4	9	1	6	2	9	3	7	6	1	1	0
0	5	5	4	6	7	7	9	6	9	0	2	5	5	3	5	8	5	1	2	9	6	9	3	9
5	7	4	3	2	8	8	4	4	2	0	8	9	6	3	0	5	1	1	2	7	3	7	8	0
8	3	2	7	1	2	7	0	2	9	1	1	7	1	5	4	8	1	9	1	2	5	0	5	3
3	1	2	1	0	7	7	3	0	4	7	1	3	8	9	3	8	7	2	7	5	1	4	8	9
0	7	9	7	0	6	4	5	3	0	5	8	2	7	3	7	3	0	6	2	4	3	3	9	1
9	0	3	4	4	3	1	8	2	1	0	4	5	9	7	2	9	0	5	5	4	7	1	5	9
1	5	7	9	2	9	5	2	8	9	1	8	6	4	2	3	4	0	6	1	4	1	7	9	9
7	3	8	2	7	8	4	7	5	9	3	4	2	9	9	4	8	3	1	1	6	5	1	5	6
2	4	0	4	4	0	4	5	0	7	6	4	9	2	0	5	3	9	2	8	1	1	8	0	2
2	9	9	9	6	6	8	0	6	9	4	0	8	4	2	4	0	4	6	0	2	1	2	2	4
5	8	2	2	2	1	7	7	2	5	9	4	2	1	7	2	1	7	7	9	3	3	5	9	8
7	3	7	4	3	6	3	0	9	9	1	6	3	9	2	3	0	2	6	8	9	8	9	0	7
8	8	9	7	6	2	9	9	0	1	2	0	0	1	0	2	4	7	8	9	6	6	9	7	8
1	4	0	9	6	1	0	9	8	7	0	5	8	0	6	5	8	0	5	0	1	9	3	0	1
1	6	4	2	4	7	6	7	7	3	5	9	3	2	2	9	2	7	8	6	3	7	7	8	1
1	2	9	8	1	2	5	7	7	9	6	8	4	4	0	6	3	3	1	1	6	7	2	5	8
5	7	7	5	3	5	5	5	6	7	9	4	3	1	5	7	2	7	6	9	7	6	1	0	3
2	4	7	9	1	7	2	8	3	4	4	1	1	1	3	0	6	9	1	4	8	8	7	5	6
0	2	5	9	4	0	8	2	5	6	0	4	7	1	6	3	6	5	5	6	1	1	6	7	6

Table 1. Random numbers (cont.)

8	9	0	8	8	8	7	4	1	9	9	9	5	5	1	8	2	1	3	7	5	7	8	7	1
1	1	0	4	2	7	2	3	9	9	5	7	5	0	9	5	3	9	6	8	6	7	4	9	0
0	0	6	6	6	3	1	5	6	3	8	9	7	2	9	0	9	8	4	9	4	2	5	0	0
2	8	5	9	9	3	5	2	5	2	1	1	7	4	0	7	9	0	1	4	9	1	9	8	9
7	5	8	0	7	9	4	5	7	9	3	2	0	7	6	3	2	6	3	6	0	9	7	8	5
2	8	1	2	4	9	9	2	0	1	9	7	9	7	2	0	8	1	4	9	2	8	6	5	9
1	6	5	9	5	2	6	8	5	8	1	8	0	6	1	2	2	7	1	0	8	6	1	9	9
3	8	0	2	2	2	0	4	5	5	5	4	5	6	9	9	1	4	2	6	7	3	9	3	5
7	0	7	8	2	1	9	6	3	1	1	8	1	1	7	8	1	6	0	3	9	6	7	1	0
9	5	9	2	6	6	6	7	4	1	9	5	1	9	8	4	2	7	9	3	8	5	5	0	8
9	9	3	7	7	0	5	3	1	2	2	4	7	0	2	2	4	0	2	1	4	5	2	6	9
2	8	6	7	5	0	2	8	7	0	4	2	5	4	1	5	3	3	7	0	7	8	8	0	8
5	8	4	6	5	0	3	6	4	5	2	4	7	9	6	7	7	3	1	5	9	7	7	4	2
2	7	9	4	0	0	1	7	0	7	2	0	0	5	1	8	6	4	9	7	9	7	0	4	8
3	2	0	4	1	5	9	2	4	0	8	3	9	0	6	9	8	3	7	7	2	6	0	6	8
9	4	4	2	4	3	1	3	1	3	0	2	2	8	2	7	5	6	8	5	3	2	9	9	9
1	4	7	7	0	3	1	3	3	5	9	6	5	1	6	4	0	6	9	7	3	9	2	1	6
2	7	4	6	7	2	6	2	7	2	5	1	3	8	7	7	8	2	1	9	2	5	0	9	0
5	3	2	1	6	4	9	4	4	6	2	5	3	3	3	5	2	5	4	9	5	7	4	4	6
6	0	9	6	4	0	0	9	3	2	7	7	6	6	7	9	7	8	1	8	0	4	1	8	1
6	8	6	5	0	5	3	4	2	3	3	7	5	7	7	9	7	4	7	0	5	6	5	1	3
7	2	1	3	4	1	7	8	1	8	4	4	1	6	6	6	2	5	6	6	2	0	4	1	9
7	5	9	1	3	2	7	1	2	6	3	1	3	3	1	2	9	0	9	8	9	8	6	9	8
8	7	7	6	8	8	8	1	6	8	6	1	8	8	6	1	7	5	6	8	6	4	3	6	9
0	4	6	4	6	1	9	6	1	4	5	9	1	1	3	6	1	4	5	7	0	8	2	5	4
9	6	8	6	1	6	3	0	3	7	0	4	9	8	8	7	7	6	8	1	7	1	5	0	8
7	6	9	7	0	9	8	7	1	2	0	9	0	3	8	5	3	9	3	7	4	1	1	5	7
3	2	7	0	9	2	7	5	8	0	4	7	8	1	4	2	4	0	0	9	6	5	9	2	5
4	2	6	8	9	1	9	0	4	2	1	3	4	3	2	0	6	7	4	7	1	3	9	7	9
6	8	6	5	1	4	1	3	0	6	7	0	9	5	2	8	7	0	9	3	8	5	1	3	5
6	3	5	7	2	0	2	8	6	3	3	8	5	3	1	0	4	6	6	3	1	7	9	9	7
7	3	7	7	3	4	5	2	3	6	2	3	6	5	5	3	9	2	1	7	0	6	4	2	0
6	0	1	2	5	0	2	9	4	9	8	3	5	9	5	7	4	5	2	8	4	7	6	6	4
2	6	6	8	6	5	0	7	7	5	5	4	9	1	2	0	3	4	8	9	6	4	9	8	9
3	6	8	7	2	9	9	2	7	5	6	0	9	0	6	5	8	8	2	8	3	4	7	4	0
4	2	5	5	7	2	6	5	9	4	3	8	7	5	6	5	3	6	3	4	3	8	5	4	7
3	2	3	1	1	5	6	5	8	3	9	6	2	2	0	2	9	0	9	3	1	1	3	1	4
0	2	3	6	6	9	4	4	6	6	0	9	9	7	4	0	1	3	2	5	6	9	4	5	1
6	5	6	9	4	1	6	8	8	6	7	0	0	6	0	8	8	3	9	7	8	4	1	7	6
7	3	1	3	9	1	2	0	7	1	5	2	1	2	0	7	0	1	7	8	6	4	6	6	3
3	5	2	5	5	9	9	0	1	5	3	2	1	7	0	1	9	3	6	3	3	4	5	0	9
2	7	6	2	3	9	6	7	5	3	6	1	5	0	2	0	3	2	9	1	6	2	1	4	6
7	8	9	1	3	0	3	0	0	2	8	5	5	4	3	8	9	6	8	2	2	1	8	8	1
1	1	0	8	2	7	9	9	8	5	5	1	9	0	7	1	2	5	7	6	8	5	8	2	8
9	6	3	9	6	2	1	1	1	0	3	2	1	7	5	0	6	9	0	6	2	0	9	5	1
1	0	3	2	4	6	1	9	9	8	8	6	5	7	6	9	8	9	1	2	4	9	1	3	5
2	3	7	1	5	7	2	5	8	1	1	7	6	6	4	9	1	3	0	3	5	2	6	3	3
2	3	6	4	7	5	3	4	7	7	7	6	4	3	5	9	6	3	8	7	8	0	1	3	2
9	3	6	1	5	4	4	5	3	3	5	4	1	5	2	3	4	6	4	5	3	7	6	9	2
0	4	0	4	6	7	0	2	9	4	3	5	9	9	7	4	9	0	6	8	7	5	9	3	6

Table 1. Random numbers (cont.)

9	3	6	4	8	6	5	9	2	6	4	5	1	6	9	9	0	8	6	7	4	5	7	2	8
1	1	5	8	8	6	9	0	3	3	6	8	4	1	8	1	3	9	0	8	3	4	5	6	5
7	2	8	1	8	8	3	7	4	4	3	5	0	2	1	3	1	9	9	1	1	1	7	0	0
1	8	4	9	4	8	6	2	6	5	1	7	6	9	5	8	8	2	8	4	0	6	2	7	8
2	7	3	0	6	1	3	6	4	1	9	2	4	5	4	4	9	5	4	7	1	4	2	0	0
2	1	0	3	9	9	3	2	8	0	0	3	4	6	2	9	2	5	5	9	6	5	0	7	8
5	1	2	1	7	3	1	5	7	1	5	8	7	7	5	7	9	8	0	8	5	3	2	5	8
2	5	3	5	4	8	4	5	2	5	7	7	2	8	7	1	8	2	3	9	3	1	5	9	9
0	6	1	5	3	1	9	8	0	4	3	2	0	1	4	5	4	2	9	8	2	9	1	5	5
4	7	0	9	2	7	5	8	6	1	5	4	0	9	9	7	3	9	6	5	5	4	0	1	4
4	6	1	4	8	5	7	1	9	7	0	9	4	2	8	0	1	3	6	4	0	4	9	7	2
8	5	2	7	5	0	5	6	6	3	3	3	1	8	1	6	7	3	2	4	9	6	6	8	9
1	9	5	1	2	4	1	4	7	2	9	8	7	7	4	9	5	1	2	8	6	7	0	0	7
1	1	7	5	2	6	4	7	5	9	2	9	2	7	0	9	3	3	1	6	2	1	0	8	2
6	0	4	0	7	7	9	9	5	0	3	8	6	9	8	9	1	2	5	2	6	3	3	6	5
4	2	8	8	4	2	2	6	5	9	7	6	4	5	2	4	4	4	7	2	3	3	8	0	1
6	3	1	3	5	0	4	8	3	4	1	7	2	9	0	6	3	3	5	0	4	0	4	5	1
4	9	9	6	2	8	3	1	8	4	8	1	1	0	9	4	6	4	2	1	5	9	4	8	6
5	5	8	5	7	3	5	3	1	0	8	9	8	0	1	0	6	2	1	6	9	7	3	5	1
0	8	3	6	4	9	7	5	6	2	8	7	3	8	9	0	2	2	0	0	4	9	9	0	9
5	6	2	1	3	3	7	4	0	7	1	9	3	8	7	6	5	8	9	0	8	3	7	1	4
6	7	6	6	5	2	7	1	5	0	1	5	8	3	1	5	3	5	5	2	2	4	2	5	4
1	0	2	9	2	0	9	5	4	1	6	9	6	8	4	0	2	6	5	3	2	2	1	3	9
9	7	3	0	4	1	8	8	6	5	9	3	9	1	2	2	0	7	2	3	8	9	9	7	8
3	6	6	7	1	6	5	6	6	9	6	7	8	6	2	1	4	1	1	0	8	8	5	4	0
2	4	3	9	7	6	0	0	6	2	8	4	3	4	4	1	1	5	9	3	7	9	4	8	3
0	4	7	0	4	1	0	7	2	9	6	4	5	2	7	2	9	8	3	4	5	6	8	8	2
6	0	5	9	1	1	1	4	4	6	9	7	8	8	6	3	6	7	6	0	5	1	0	5	5
1	1	5	1	6	6	0	5	1	5	6	0	7	5	2	7	3	7	2	4	8	6	2	5	4
3	4	2	3	2	5	9	4	7	1	7	8	4	1	3	8	8	5	3	7	6	8	8	6	4
8	3	3	6	5	8	0	5	9	6	6	1	3	4	5	4	2	8	3	9	5	0	8	9	1
9	2	1	2	4	7	6	5	9	3	6	0	5	0	7	5	3	7	9	3	8	5	1	7	6
2	6	6	8	4	7	5	4	7	0	8	4	2	6	8	3	1	4	5	9	8	7	5	0	6
6	6	4	6	5	8	8	5	9	8	5	9	4	6	5	2	4	0	7	1	4	1	8	7	0
1	1	6	5	4	5	4	0	4	1	7	2	1	5	7	5	8	5	7	4	4	8	2	6	2
3	0	8	3	7	1	3	1	9	0	7	7	5	2	2	7	6	3	9	9	9	0	3	8	6
8	0	2	6	1	8	5	9	3	1	7	9	4	7	5	5	4	9	6	4	6	1	6	0	1
4	5	2	7	5	1	0	6	4	2	1	6	2	4	9	1	8	3	1	8	8	2	7	4	1
0	5	6	1	3	8	3	9	8	3	6	9	4	9	1	5	2	5	6	5	8	4	5	1	9
7	4	1	5	0	4	4	3	4	8	7	4	8	7	4	5	1	3	9	2	4	1	2	2	5
7	4	5	7	0	9	8	3	4	9	7	8	1	3	2	2	8	3	7	3	8	5	2	6	1
5	8	8	2	4	5	4	9	5	6	5	5	0	1	7	6	3	6	1	6	6	5	6	8	9
1	4	9	9	2	0	5	4	1	2	6	4	3	8	4	3	4	3	2	4	4	2	9	5	6
2	3	5	4	3	3	6	9	2	8	2	1	1	5	5	0	7	1	4	5	0	5	6	3	0
9	6	1	5	9	9	1	2	9	2	5	3	9	9	4	1	6	2	3	4	0	8	8	6	9
0	7	2	9	3	7	5	5	5	0	5	7	3	3	6	8	6	2	7	2	1	5	0	0	3
6	2	8	1	5	1	1	4	8	2	9	5	5	6	5	2	0	6	7	3	3	9	2	2	2
2	7	8	8	9	0	4	1	4	6	9	7	5	4	9	2	4	4	0	6	9	5	4	4	4
4	3	3	9	1	2	1	3	6	3	4	3	4	8	8	6	9	3	2	3	3	4	7	1	2
8	8	0	5	2	2	8	0	8	5	3	0	3	7	4	9	6	0	1	8	5	3	8	6	4

Source: This table was generated by James B. Weaver, III.

Table 2. Distribution of t

df	Level of significance for one-tailed test					
	.10	.05	.025	.01	.005	.0005
	Level of significance for two-tailed test					
	.20	.10	.05	.02	.01	.001
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.965	9.925	31.598
3	1.638	2.353	3.182	4.541	5.841	12.941
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.859
6	1.440	1.943	2.447	3.143	3.707	5.959
7	1.415	1.895	2.365	2.998	3.499	5.405
8	1.397	1.860	2.306	2.896	3.355	5.041
9	1.383	1.833	2.262	2.821	3.250	4.781
10	1.372	1.812	2.228	2.764	3.169	4.587
11	1.363	1.796	2.201	2.718	3.106	4.437
12	1.356	1.782	2.179	2.681	3.055	4.318
13	1.350	1.771	2.160	2.650	3.012	4.221
14	1.345	1.761	2.145	2.624	2.977	4.140
15	1.341	1.753	2.131	2.602	2.947	4.073
16	1.337	1.746	2.120	2.583	2.921	4.015
17	1.333	1.740	2.110	2.567	2.898	3.965
18	1.330	1.734	2.101	2.552	2.878	3.992
19	1.328	1.729	2.093	2.539	2.861	3.883
20	1.325	1.725	2.086	2.528	2.845	3.850
21	1.323	1.721	2.080	2.518	2.831	3.819
22	1.321	1.717	2.074	2.508	2.819	3.792
23	1.319	1.714	2.069	2.500	2.807	3.767
24	1.318	1.711	2.064	2.492	2.797	3.745
25	1.316	1.708	2.060	2.485	2.787	3.725
26	1.315	1.706	2.056	2.479	2.779	3.707
27	1.314	1.703	2.052	2.473	2.771	3.690
28	1.313	1.701	2.048	2.467	2.763	3.674
29	1.311	1.699	2.045	2.462	2.756	3.659
30	1.310	1.697	2.042	2.457	2.750	3.646
40	1.303	1.684	2.021	2.423	2.704	3.551
60	1.296	1.671	2.000	2.390	2.660	3.460
120	1.289	1.658	1.980	2.358	2.617	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.291

*This table is abridged from Table III of R. A. Fisher and F. Yates *Statistical Tables for Biological, Agricultural, and Medical Research*, published by Oliver and Boyd, Ltd., Edinburgh, by permission of the authors and publishers.

Table 3. Areas under the normal curve (Proportion of area under the normal curve between the mean and a *z* Distance from the mean)

$\frac{x}{\sigma}$ or z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998
3.5	.4998									
4.0	.49997									
4.5	.499997									
5.0	.4999997									

Table 4. Distribution of chi-square

	Probability					
df	.20	.10	.05	.02	.01	.001
1	1.642	2.706	3.841	5.412	6.635	10.827
2	3.219	4.605	5.991	7.824	9.210	13.815
3	4.642	6.251	7.815	9.837	11.345	16.266
4	5.989	7.779	9.488	11.668	13.277	18.467
5	7.289	9.236	11.070	13.388	15.086	20.515
6	8.558	10.645	12.592	15.033	16.812	22.457
7	9.803	12.017	14.067	16.622	18.475	24.322
8	11.030	13.362	15.507	18.168	20.090	26.125
9	12.242	14.684	16.919	19.679	21.666	27.877
10	13.442	15.987	18.307	21.161	23.209	29.588
11	14.631	17.275	19.675	22.618	24.725	31.264
12	15.812	18.549	21.026	24.054	26.217	32.909
13	16.985	19.812	22.362	25.472	27.688	34.528
14	18.151	21.064	23.685	26.873	29.141	36.123
15	19.311	22.307	24.996	28.259	30.578	37.697
16	20.465	23.542	26.296	29.633	32.000	39.252
17	21.615	24.769	27.587	30.995	33.409	40.790
18	22.760	25.989	28.869	32.346	34.805	42.312
19	23.900	27.204	30.144	33.687	36.191	43.820
20	25.038	28.412	31.410	35.020	37.566	45.315
21	26.171	29.615	32.671	36.343	38.932	46.797
22	27.301	30.813	33.924	37.659	40.289	48.268
23	28.429	32.007	35.172	38.968	41.638	49.728
24	29.553	33.196	36.415	40.270	42.980	51.179
25	30.675	34.382	37.652	41.566	44.314	52.620

Table 4. (cont'd) Distribution of chi-square

	Probability					
df	.20	.10	.05	.02	.01	.001
26	31.795	35.563	38.885	42.856	45.642	54.052
27	32.912	36.741	40.113	44.140	46.963	55.476
28	34.027	37.916	41.337	45.419	48.278	56.893
29	35.139	39.087	42.557	46.693	49.588	58.302
30	36.250	40.256	43.773	47.962	50.892	59.703
32	38.466	42.585	46.194	50.487	53.486	62.487
34	40.676	44.903	48.602	52.995	56.061	65.247
36	42.879	47.212	50.999	55.489	58.619	67.985
38	45.076	49.513	53.384	57.969	61.162	70.703
40	47.269	51.805	55.759	60.436	63.691	73.402
42	49.456	54.090	58.124	62.892	66.206	76.084
44	51.639	56.369	60.481	65.337	68.710	78.750
46	53.818	58.641	62.830	67.771	71.201	81.400
48	55.993	60.907	65.171	70.197	73.683	84.037
50	58.164	63.167	67.505	72.613	76.154	86.661
52	60.332	65.422	69.832	75.021	78.616	89.272
54	62.496	67.673	72.153	77.422	81.069	91.872
56	64.658	69.919	74.468	79.815	83.513	94.461
58	66.816	72.160	76.778	82.201	85.950	97.039
60	68.972	74.397	79.082	84.580	88.379	99.607
62	71.125	76.630	81.381	86.953	90.802	102.166
64	73.276	78.860	83.675	89.320	93.217	104.716
66	75.424	81.085	85.965	91.681	95.626	107.258
68	77.571	83.308	88.250	94.037	98.028	109.791
70	79.715	85.527	90.531	96.388	100.425	112.317

This table is adapted from R. A. Fisher and F. Yates, *Statistical Tables for Biological, Agricultural and Medical Research*, Oliver and Boyd, Ltd., Edinburgh, by permission of the authors and publishers.

Table 5. Distribution of F
(.05 Level)

df_1 \ df_2	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88

20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

This table is abridged from Table 18 of the *Biometrika Tables for Statisticians*, Vol. 1(ed. 1), edited by E. S. Pearson and H. O. Hartley. Reproduced by the kind permission of E. S. Pearson and the trustees of *Biometrika*.

Table 5 (cont'd) Distribution of F
(.01 Level)

df_1 \ df_2	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	4052	4999.5	5403	5625	5764	5859	5928	5982	6022	6056	6106	6157	6209	6235	6261	6287	6313	6339	6366
2	98.5	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.49	99.50
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.05	26.87	26.69	26.60	26.50	26.41	26.32	26.22	26.13
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.81	6.72	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49

20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	2.96	2.81	2.66	2.58	2.50	2.42	2.33	2.23	2.13
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	2.10
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00	2.87	2.73	2.57	2.49	2.41	2.33	2.23	2.14	2.03
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00

APPENDIX 2

Computer Program for Random Telephone Number Generation (TELEGEN) TELEGEN Example

```
 82/02/19. 12.07.08.
PROGRAM     TELEGEN

00070  REM  TELEGEN WAS WRITTEN ON A CDC CYBER-730 SYSTEM WITH NO
00080  REM  MACHINE ORIENTED STATEMENTS. IT SHOULD BE EASILY ADAPTABLE
00090  REM  TO ANY COMPUTER (APPLE, ETC.).
00100  RANDOMIZE
00101  PRINT  ,,"TELEGEN"
00102  PRINT
00103  PRINT  ,"          JAMES B. WEAVER, III"
00104  PRINT
00105  PRINT
00110  PRINT  "       TELEGEN IS DESIGNED TO GENERATE RANDOM TELEPHONE"
00120  PRINT  "NUMBERS BASED ON A SET OF THREE DIGIT LOCAL EXCHANGE"
00130  PRINT  "PREFIX NUMBERS FURNISHED BY THE USER. YOU ARE ALLOWED A"
00140  PRINT  "MAXIMUM OF 10 LOCAL PREFIX ENTRIES. THE NUMBER OF RANDOM"
00150  PRINT  "TELEPHONE NUMBERS TO BE GENERATED PER"
00155  PRINT  "EXCHANGE MUST BE ENTERED IN MULTIPLES OF TWO"
00160  PRINT  "(2, 4, 6, ETC.)."
00170  PRINT
00180  PRINT  "       TELEGEN IS EASY TO USE. FOR EXAMPLE ... IF YOU WANT TO"
00190  PRINT  "GENERATE -2- RANDOM TELEPHONE NUMBERS FOR THE LOCAL"
00200  PRINT  "EXCHANGES -100- AND -200- YOU WOULD ENTER YOUR REQUEST"
00210  PRINT  "AS FOLLOWS..."
00220  PRINT
00230  PRINT  "ENTER LOCAL EXCHANGE PREFIXES SEPARATED BY COMMAS (10"
00240  PRINT  "MAX.) NOTE ... 10 NUMERIC ENTRIES MUST BE MADE. FOR ENTRIES"
00250  PRINT  "WITHOUT           THREE DIGIT EXCHANGE NUMBERS"
00260  PRINT  "ENTER '0'. ?100,200,0,0,0,0,0,0,0,0"
00270  PRINT
00280  PRINT  "ENTER THE NUMBER OF RANDOM TELEPHONE NUMBERS TO BE"
00290  PRINT  "GENERATED FOR EACH EXCHANGE. MUST BE A MULTIPLE OF TWO."
00300  PRINT  "? 2"
00310  PRINT
00320  PRINT  "YOUR RESULTS WOULD BE THE LOCAL PREFIX PLUS FOUR"
00330  PRINT  "RANDOMLY SELECTED DIGITS. FOR EXAMPLE ... 353-0123"
00340  PRINT  "                                353-4567   542-8901   542-2345"
00350  PRINT
00360  PRINT  "NOTE ..."
00370  PRINT  "       THIS PROGRAM WILL PRINT ALL RANDOM TELEPHONE"
00380  PRINT  "NUMBERS FOR EACH LOCAL EXCHANGE TOGETHER TO ALLOW"
00390  PRINT  "YOU TO CHECK FOR POSSIBLE DUPLICATION."
00400  PRINT
00410  PRINT
00420  PRINT  "WHEN YOU ARE READY TO BEGIN THIS PROGRAM ENTER A 'Y'."
00430  INPUT  A$
00440  IF A$="Y" GOTO 00450
00450  PRINT
00460  PRINT  "ENTER LOCAL EXCHANGE PREFIXES SEPARATED BY COMMAS (10"
00470  PRINT  "MAX.) NOTE ... 10 NUMERIC ENTRIES MUST BE MADE. FOR ENTRIES"
00480  PRINT  "WITHOUT           THREE DIGIT EXCHANGE NUMBERS ENTER '0'."
00490  INPUT  A,B,C,D,E,F,G,H,I,J
00500  PRINT
00510  PRINT  "ENTER THE NUMBER OF RANDOM TELEPHONE NUMBERS TO BE"
00520  PRINT  "GENERATED FOR EACH EXCHANGE. MUST BE A MULTIPLE OF TWO."
00530  INPUT  Z
00540  IF A>0 THEN A1=1
00550  IF B>0 THEN B1=1
```

```
00560 IF C>0 THEN C1=1
00570 IF D>0 THEN D1=1
00580 IF E>0 THEN E1=1
00590 IF F>0 THEN F1=1
00600 IF G>0 THEN G1=1
00610 IF H>0 THEN H1=1
00620 IF I>0 THEN I1=1
00630 IF J>0 THEN J1=1
00640 Y=A1+B1+C1+D1+E1+F1+G1+H1+I1+J1
00650 IF Y=0 THEN GOTO 5000
00660 IF Z=0 THEN GOTO 5010
00665 PRINT
00670 X=Y*Z
00675 PRINT
00680 PRINT  Z; "RANDOM TELEPHONE NUMBERS WILL BE GENERATEO FOR EACH"
00690 PRINT  "OF THE ";Y;" LOCAL EXCHANGES THAT YOU HAVE ENTERED. THIS"
00700 PRINT  "WILL RESULT IN A TOTAL OF ";X;" RANDOM TELEPHONE NUMBERS."
00720 PRINT
00725 Z1=Z/2
00730 PRINT
00740 IF A=0 THEN 5000
00750 PRINT ,"FOR THE ";A;"- EXCHANGE"
00760 PRINT
00770 FOR W=1 TO Z1
00780 GOSUB 4000
00790 PRINT "        "A" - "K;L;M;N,A" - "K1;L1;M1;N1
00800 NEXT W
00810 PRINT
00818 PRINT
00819 IF B=0 THEN 5020
00820 PRINT ,"FOR THE "B;"- EXCHANGE"
00830 PRINT
00840 FOR W=1 TO Z1
00850 GOSUB 4000
00860 PRINT "        "B" - "K;L;M;N,B" - "K1;L1;M;N1
00870 NEXT W
00880 PRINT
00890 PRINT
00900 IF C=0 THEN 5020
00910 PRINT ,"FOR THE";C;"- EXCHANGE"
00920 PRINT
00930 FOR W=1 TO Z1
00940 GOSUB 4000
00950 PRINT "        "C" - "K;L;M;N,B" - "K1;L1;M1;N1
00960 NEXT W
00970 PRINT
00980 PRINT
00990 IF D=0 THEN 5020
01000 PRINT ,"FOR THE";D;- EXCHANGE
01010 PRINT
01020 FOR W=1 TO Z1
01030 GOSUB 4000
01040 PRINT "        "D" - "K;L;M;N,D" - "K1;L1;M1;N1
01050 NEXT W
01060 PRINT
01070 PRINT
01080 IF E=0 THEN 5020
01090 PRINT ,"FOR THE ";E;"- EXCHANGE"
01100 PRINT
01110 FOR W=1 TO Z1
```

```
01120 GOSUB 4000
01130 PRINT "          "E" - "K;L;M;N,E" - "K1;L;M1N1
01140 NEXT W
01150 PRINT
01160 PRINT
01170 IF F=0 THEN 5020
01180 PRINT ,"FOR THE";F;"- EXCHANGE
01190 PRINT
01200 FOR W=1 TO Z1
01210 GOSUB 4000
01220 PRINT "          "F" - "K;L;M;N,F" - "K1;L1;M1;N1
01230 NEXT W
01240 PRINT
01250 PRINT
01260 IF G= 0 THEN 5020
01270 PRINT ,"FOR THE";G;"- EXCHANGE
01280 PRINT
01290 FOR W=1 TO Z1
01300 GOSUB 4000
01310 PRINT "          "G" - "K;L;M;N,G" - "K1;L1;M1;N1
01320 NEXT W
01330 PRINT
01340 PRINT
01350 IF H=0 THEN 5020
01360 PRINT ,"FOR THE";H;"- EXCHANGE"
01370 PRINT
01380 FOR W=1 TO Z1
01390 GOSUB 4000
01400 PRINT "          "H" - "K;L;M;N,H" - "K1;L1;M1;N1
01410 NEXT W
01420 PRINT
01430 PRINT
01440 IF I=0 THEN 5020
01450 PRINT, "FOR THE";I;- EXCHANGE"
01460 PRINT
01470 FOR W=1 TO Z1
01480 GOSUB 4000
01490 PRINT "          "I" - "K;L;M;N,I" - "K1;L1;M1;N1
01500 NEXT W
01510 PRINT
01520 PRINT
01530 IF J=0 THEN 5020
01540 PRINT ,"FOR THE";J;"- EXCHANGE"
01550 PRINT
01560 FOR W=1 TO Z1
01570 GOSUB 4000
01580 PRINT "          "J" - "K;L;M;N,J" - "K1;L1;M1;N1
01590 NEXT W
01600 PRINT
01610 PRINT
01620 GOTO 5020
3999 REM RANDOM NUMBER GENERATOR SUBROUTINE
4000 K=INT(10*RND(0))
4010 L=INT(10*RND(0))
4020 M=INT(10*RND(0))
4030 N=INT(10*RND(0))
4040 K1=INT(10*RND(0))
4050 L1=INT(10*RND(0))
4060 M1=INT(10*RND(0))
4070 N1=INT (10*RND(0))
```

```
4080 RETURN
5000 PRINT "NO PREFIX ENTERED. PROGRAM TERMINATED."
5001 GOTO 5020
5010 PRINT  "NO RANDOM NUMBERS REQUESTED. PROGRAM TERMINATED."
5011 GOTO 5020
5020 PRINT  "END OF PROGRAM"
5030 END
```

TELEGEN EXAMPLE

TELEGEN IS DESIGNED TO GENERATE RANDOM TELEPHONE NUMBERS BASED ON A SET OF THREE DIGIT LOCAL EXCHANGE PREFIX NUMBERS FURNISHED BY THE USER. YOU ARE ALLOWED A MAXIMUM OF 10 LOCAL PREFIX ENTRIES. THE NUMBER OF RANDOM TELEPHONE NUMBERS TO BE GENERATED PER EXCHANGE MUST BE ENTERED IN MULTIPLES OF TWO (2,4,6, ETC.).

TELEGEN IS EASY TO USE. FOR EXAMPLE ... IF YOU WANT TO GENERATE -2- RANDOM TELEPHONE NUMBERS FOR THE LOCAL EXCHANGES -100- AND -200- YOU WOULD ENTER YOUR REQUEST AS FOLLOWS ...

```
ENTER LOCAL EXCHANGE PREFIXES SEPARATED BY COMMAS (10 MAX.).
NOTE ... 10 NUMERIC ENTRIES MUST BE MADE. FOR ENTRIES WITHOUT THREE
          DIGIT EXCHANGE NUMBERS ENTER '0'.
? 100,200,0,0,0,0,0,0,0,0

ENTER THE NUMBER OF RANDOM TELEPHONE NUMBERS TO BE GENERATED FOR
EACH EXCHANGE. MUST BE A MULTIPLE OF TWO.

YOUR RESULTS WOULD BE THE LOCAL PREFIX PLUS FOUR RANDOMLY SELECTED
DIGITS. FOR EXAMPLE ... 353-0123  353-4567
                        542-8901  542-2345

NOTE ...
      THIS PROGRAM WILL PRINT ALL RANDOM TELEPHONE NUMBERS FOR EACH
LOCAL EXCHANGE TOGETHER TO ALLOW YOU TO CHECK FOR POSSIBLE
DUPLICATION.


WHEN YOU ARE READY TO BEGIN THIS PROGRAM ENTER A 'Y'.
? y


ENTER LOCAL EXCHANGE PREFIXES SEPARATED BY COMMAS (10 MAX).
NOTE ... 10 NUMERIC ENTRIES MUST BE MADE. FOR ENTRIES WITHOUT THREE
          DIGIT EXCHANGE NUMBERS ENTER '0'.
? 100,200,300,400,500,600,700,800,900,999

ENTER THE NUMBER OF RANDOM TELEPHONE NUMBERS TO BE GENERATED FOR
EACH EXCHANGE. MUST BE A MULTIPLE OF TWO.


2 RANDOM TELEPHONE NUMBERS WILL BE GENERATED FOR EACH OF THE 10
LOCAL EXCHANGES THAT YOU HAVE ENTERED. THIS WILL RESULT IN A TOTAL OF
20 RANDOM TELEPHONE NUMBERS.
```

```
        FOR THE 100 - EXCHANGE

100 - 3 1 4 0       100 - 0 5 1 8

        FOR THE 200 - EXCHANGE

200 - 4 6 4 0       200 - 5 6 3 3

        FOR THE 300 - EXCHANGE

300 - 8 4 7 2       300 - 7 5 4 0

        FOR THE 400 - EXCHANGE

400 - 1 1 1 9       400 - 8 0 7 8

        FOR THE 500 - EXCHANGE

500 - 5 6 6 5       500 - 6 9 9 3

        FOR THE 600 - EXCHANGE

600 - 7 9 5 3       600 - 7 0 8 9

        FOR THE 700 - EXCHANGE

700 - 3 9 1 1       700 - 5 0 7 7

        FOR THE 800 - EXCHANGE

800 - 9 7 4 5       800 - 2 2 7 3

        FOR THE 900 - EXCHANGE

900 - 0 4 7 7       900 - 7 5 6 6

        FOR THE 999 - EXCHANGE

999 - 8 0 9 4       999 - 6 4 3 5

ENO OF PROGRAM

SRU      0.683 UNTS.

RUN COMPLETE.
```

Glossary

accretion measure: an observation of those patterns of behavior or activities that leave behind some physical trace (usually used in connection with unobtrusive research)

acoustical coupler: a device used to allow remote access to a main frame computer via a telephone receiver hook-up

advertising element research: a type of message research that involves testing each element of an advertisement (for example, headline, illustrations, typography) for effectiveness

aided recall: a survey technique in which the interviewer shows the respondent a copy of a newspaper, magazine, television schedule, or other item that might help him or her to remember a certain article, program, advertisement, and so on

analysis of variance (ANOVA): a statistical procedure used to decompose sources of variation in two or more independent variables

analytical survey: a survey that attempts to describe and explain why certain conditions exist (usually by testing certain hypotheses)

antecedent variable: (1) in survey research, the variable used to predict another variable; (2) in experimental research, the independent variable

applied research: research that attempts to solve a specific problem rather than to construct a theory

area of dominant influence (ADI): a region composed of a certain number of television households; every county is assigned to one and only one ADI

artifact: a variable that creates a rival explanation of results (a confounding variable)

audience turnover: in radio research, an estimate of the number of times the audience changes stations during a given daypart

audimeter: an electronic audience measurement device used by the A. C. Nielsen Company to collect data

available sample: a sample selected on the basis of accessibility

average quarter hour: the number of persons or households tuned in to a specific channel or station for at least 5 minutes during a 15-minute time segment

beta weight: a mathematically derived value representing a variable's contribution to a prediction or weighted linear combination (also called *weight coefficient*)

call-out research: a procedure used in radio research to determine the popularity of recordings

canned computer program: a ready-to-use statistical package containing a variety of statistical subprograms

canonical correlation: a multivariate statistic used to investigate the relationship between two sets of variables

census: an analysis in which the sample comprises every element of a population

central memory: the part of the computer where data are temporarily stored in order to solve specific problems (also known as *core storage*)

central processing unit (cpu): the control and coordination center of a computer

central tendency: a single value that is chosen to represent a typical score in a distribution such as the mean, mode, and median

checklist: a type of questionnaire in which the respondent is given a list of items and is asked to check those that apply

chi-square statistic: a measurement of observed frequencies vs. expected frequencies

circulation: in the print media, the total number of copies of a newspaper or magazine that are delivered to a subscriber plus all copies bought at newsstands or from other sellers

circulation research: (1) a market-level study of newspaper and magazine penetration; (2) a study of the delivery and pricing systems used by newspapers and magazines

close-ended question: a type of questionnaire in which the respondent must select his or her answer from a prepared set of options

cloze procedure: a method for measuring readability or recall in which every *n*th word is deleted from the message and readers are asked to fill in the blanks

cluster sample: a sample selected in groups or categories

coding: in content analysis, placing a unit of analysis into a particular category

coefficient of determination: in correlational statistics, the amount of variation in the criterion variable that is accounted for by the antecedent variable

coefficient of nondetermination: in correlational statistics, the amount of variation in the criterion variable that is left unexplained

communication audit: in public relations, an examination of the internal and external means of communication used by an organization

complete execution: a form of testing in which an entire advertisement or commercial is used as a stimulus

computer cards: 3¼″ × 7⅜″ cards used to input data into a computer

concept: a term that expresses an abstract idea formed by generalization

confidence interval: an area within which there is a stated probability that the parameter will fall

constitutive definition: a type of definition in which other words or concepts are substituted for the word being defined

construct: a combination of concepts that is created for a specific situation (for example, "authoritarianism")

constructive replication: an analysis of a hypothesis taken from a previous study that deliberately avoids duplication of the methods used in the previous study

continuous variable: a variable that can take on any value over a range of values and can be meaningfully broken into smaller subparts (for example, "height")

control group: subjects who do not receive experimental treatment and thus serve as a basis of comparison in an experiment

control variable: a variable whose influence a researcher wishes to eliminate

copy testing: see *message research*

cost per thousand (CPM): the amount of money it costs to reach 1,000 people or households by means of a particular medium or advertising vehicle

criterion variable: (1) in survey research, the variable presumed to be the effects variable; (2) in experimental research, the dependent variable

cross-lagged study: a type of longitudinal study in which information about two different variables is gathered from the same sample at two different times. The correlations between variables at the same point in time are compared with the correlations at different points in time

cross-tabulation analysis (crosstabs): see *chi-square statistic*

cross-validation: a procedure whereby measurement instruments or subjects' responses are cross-checked to verify their validity or truthfulness

cultivation analysis: a research approach that suggests that heavy television viewing leads to perceptions of social reality that are consistent with the view of the world as presented on television

cume: an estimate of the number of different persons who listened to or viewed a particular broadcast for at least five minutes during a given daypart

data archives: data storage facilities where researchers can deposit data for other researchers to use

daypart: a given part of the broadcast day (for example, prime time = 8:00–11:00 EST)

demand characteristic: the premise that subjects' awareness of the experimental condition may affect their performance in the experiment (also known as the *Hawthorne effect*)

dependent variable: the variable that is observed and whose value is presumed to depend on the independent variable(s)

descriptive statistics: those statistical methods and techniques designed to reduce data sets to allow for easier interpretation

descriptive survey: a survey that attempts to picture or document current conditions or attitudes

design-specific results: research results that are based on, or specific to, the research design used

determinant: a scalar that represents a unique function of the numbers in a square matrix

diagonal matrix: a square matrix whose elements all equal zero, except for those along the principal diagonal

discrete variable: a variable that can be conceptually subdivided into a finite number of indivisible parts (for example, the number of children in a family)

discriminant analysis: a multivariate statistic used to classify groups according to variable similarities or to analyze the statistical significance of a weighted linear combination of variables

dispersion: the amount of variability in a set of scores

distribution: a collection of scores or measurements

double-barreled question: a single question that in reality requires two separate responses (for example, "Do you like the price and style of this item?")

double blind experiment: a research study where experimenters and others involved in the study do not know whether a given subject belongs to the experimental or the control group

dummy variable: the variable created when a variable at the nominal level is transformed into a form more appropriate for higher order statistics

editor–reader comparison: a readership study in which the perceptions of editors and readers are compared

environmental monitoring program: in public relations research, a study of trends in public opinion and events in the social environment that may have a significant impact on an organization

erosion measure: any measure based on observations of wear and tear on objects resulting from human behavior (usually used in connection with unobtrusive research)

error variance: error created by an unknown factor

evaluation apprehension: a fear of being measured or tested that may result in invalid data

exhaustivity: a state of a category system such that every unit of analysis can be placed into an existing slot

experimental design: a blueprint or set of plans for conducting laboratory research

external validity: the degree to which the results of a research study are generalizable to other situations

factor analysis: a multivariate statistical procedure used primarily for data reduction, construct development, and the investigation of variable relationships

factorial design: a simultaneous analysis of two or more independent variables or factors

factor score: a composite or summary score produced by factor analysis

feeling thermometer: a rating scale patterned after a normal thermometer on which respondents can rate their attitudes on a scale of 0 to 100

filter question: a question designed to screen out certain individuals from further participation in a study

Flesch Reading Ease Formula: an early readability formula based on the number of words per sentence and the number of syllables per word

focus group: an interview conducted with two or more subjects simultaneously and a moderator who leads a discussion about a specific topic

Fog Index: a readability scale based on sentence length and the number of syllables per word

forced choice question: a question that requires a subject to choose between two specified responses

frequency curve: a graphical display of frequency data in the form of a smooth, unbroken curve

frequency distribution: a collection of scores, ordered according to magnitude, and their respective frequencies

frequency polygon: a series of lines connecting points that represent the frequencies of scores

gross rating points: the total of audience ratings during two or more time periods, representing the size of the gross audience

histogram: a bar chart that illustrates frequencies and scores

homogeneity: equality of control and experimental groups prior to an experiment (also called *point of prior equivalency*)

hook: a short representative sample of a recording used in call-out research

hypothesis: a tentative generalization concerning the relationship between two or more variables that predicts an experimental outcome

identity matrix: a square matrix whose elements equal zero except for those along the principal diagonal, which equal one

incidence: the percentage of a population that possesses the desired characteristics for a particular research study

independent variable: the variable that is systematically varied by the researcher

inferential statistics: statistical methods and techniques that allow a researcher to generalize sample results to a population within a given margin of probable error

instrumental replication: a study that duplicates the dependent variable of a previous study

instrument decay: the deterioration of a measurement instrument during the course of a study, thereby reducing the instrument's effectiveness and accuracy

interaction: a treatment-related effect dependent upon a concomitant influence of two independent variables on a dependent variable

intercoder reliability: in content analysis, the degree of agreement between or among independent coders

intercorrelation matrix: a matrix composed of correlations between pairs of variables

internal validity: a property of a research study such that results are based on expected conditions rather than on extraneous variables

interval level: a measurement system in which the intervals between adjacent points on a scale are equal (for example, a thermometer)

isomorphism: similarity of form or structure

item pretest: a method of testing subjects' interest in reading magazine or newspaper articles

item-selection study: a readership study used to determine who reads specific parts of a newspaper

kurtosis: the degree of "peakedness" in a frequency curve (curves can be *leptokurtic*—large number of scores around the center; *platykurtic*—spread over a wide area; or *mesokurtic*—two or more areas having a high concentration of scores)

leading question: a question that suggests a certain response or makes an implicit assumption (for example, "How long have you been an alcoholic?")

literal replication: a study that is an exact duplication of a previous study

mailing list: a compilation of names and addresses, sometimes prepared by a commercial firm, that is used as a sampling frame for mail surveys

mail survey: a research method whereby self-administered questionnaires are mailed to a sample of people whom the research must rely upon to mail back their responses

main effect: the effect of the independent variable(s) on the dependent variable (no interaction is present)

marker variable: a variable that highlights or defines the construct under study

masked recall: a survey technique in which the interviewer shows respondents the front cover of a newspaper or magazine with the name of the publication blacked out

matrix: a two-dimensional array of scalars

mean: the arithmetic average of a set of scores

measurement: a procedure whereby a researcher assigns numerals to objects, events, or properties according to certain rules

measurement error: an inconsistency produced by the instruments used in a research study

media efficiency: reaching the maximum possible audience for the smallest possible cost

median: the midpoint of a distribution of scores

medium variables: in a content analysis, those aspects of content that are unique to the medium under consideration (for example, typography to a newspaper or magazine)

message research: research used to determine the most effective way of structuring a message in order to achieve desired results (also known as *copy testing*)

method of authority: a method of knowing whereby something is believed because a source seen as an authority says it is true

method of intuition: a method of knowing whereby something is believed because it is "self-evident" or "stands to reason" (also called *a priori reasoning*)

method of tenacity: a method of knowing whereby something is believed because a person has always believed it to be true

method-specific results: research results based on, or specific to, the research method used

metro survey area: a broadcasting region representing one of the Standard Metropolitan Statistical Areas (SMSA), as defined by the U.S. Office of Management and Budget

mode: the score that occurs most often in a frequency distribution

mortality: a problem in panel studies and other forms of longitudinal research caused by original sample members who drop out of the research project for one reason or another

multiple regression: an analysis of two or more independent variables and their relationship to a single dependent variable (used to predict the dependent variable)

multivariate analysis of variance (MANOVA): a multivariate extension of analysis of variance used to study more than one dependent variable

multivariate statistics: statistical methods that investigate the relationship between one or more independent variables and more than one dependent variable

mutually exclusive: a state of a category system such that a unit of analysis can be placed in one and only one category

nominal level: the level of measurement at which arbitrary numerals or other symbols are used to classify persons, objects, or characteristics

nonparametric statistics: statistical procedures used with variables measured at the nominal or ordinal level

non-probability sample: a sample selected without regard to the laws of mathematical probability

normal curve: a symmetrical, bell-shaped curve that possesses specific mathematical characteristics

null hypothesis: the denial or negation of a research hypothesis

objective function: a mathematical formula that provides various quantitative values for a given media schedule of advertisements. Used in computer simulations of advertising media schedules

open-ended question: a question to which respondents are asked to provide their own answer(s) (for example, "What is your favorite type of television program?")

operational definition: a definition that specifies patterns of behavior and procedures in order to experience or measure a concept

operational replication: a study that duplicates only the sampling methodology and experimental procedures of a previous study

ordinal level: the level of measurement at which items are ranked along some continuum

overnights: ratings surveys of a night's television viewing computed in five major U.S. cities by the A. C. Nielsen Company

panel study: a research technique whereby the same sample of respondents is measured at different points in time

parameter: a characteristic or property of a population

parametric statistics: statistical procedures appropriate with variables measured at the interval or ratio level

parsimony principle: the premise that the simplest method is the most preferable (also known as *Occam's razor*)

partial correlation: a method used to control a confounding or spurious variable that may affect the relationship between independent and dependent variables

path analysis: a multivariate statistic used to create a "causal path" for dependent variables and their relationships to independent variables

periodicity: any form of bias resulting from the use of a nonrandom list of subjects or items in selecting a sample

personal interview: a survey technique whereby a trained interviewer visits the respondent and administers the questionnaire in a face-to-face setting

pilot study: a trial run of a study conducted on a small scale to determine if the research design and methodology are relevant and effective

population: a group or class of objects, subjects, or units

population distribution: the frequency distribution of all the variables of interest as determined by a census of the population

power: the probability of rejecting the null hypothesis when an alternative is true

predictor variable: see *antecedent variable*

prestige bias: the tendency of a respondent to give answers that will make him or her seem more educated, successful, financially stable, and so on

principle of least squares: a mathematical procedure used to determine a prediction line or vector through a set of data points

probability level: a predetermined value at which researchers test their data for statistical significance

probability sample: a sample selected according to the laws of mathematical probability

proposition: a statement of the form "if A then B" which links two or more concepts

proprietary data: research data gathered by a private organization that are not available to the general public unless released by that organization

psychographics: an area of research that examines why people behave and think as they do

public relations audit: a comprehensive study of the public relations position of an organization

purposive sample: a sample which is purposely chosen to be representative of a population

qualitative research: a description or analysis of a phenomenon that does not involve specific measurements of variables

quantitative research: a description or analysis of a phenomenon that does involve specific measurements of variables

quota sample: a sample selected to represent certain characteristics of interest

random digit dialing: a random method of selecting telephone numbers which insures that all telephone households have an equal chance of being selected

random error: error in a research study that cannot be controlled by the researcher

random sample: a subgroup or subset of a population selected in such a way that each unit in a population has an equal chance of being selected

range: a measure of dispersion based on the difference between the highest and lowest scores in a distribution

rating: an estimate of the percentage of people or households in a population that are tuned to a specific station or network

ratio level: a level of measurement that has all the properties of an interval level scale and also has a true zero point

reactivity: a subject's awareness of being measured or observed and its possible impact on that subject's behavior

readability: the total of all elements within a piece of printed material that affect the degree to which people understand the piece and find it interesting

reader–non-reader study: a study that contrasts non-readers of newspapers or magazines with regular readers

reader profile: a demographic summary of the readers of a particular publication

recall study: a study in which respondents are asked to remember what advertisements they remember seeing in various media

recognition: a measurement of readership whereby respondents are shown the logo of a magazine or newspaper

redundancy index: a mathematical procedure used in canonical correlation to aid in interpreting relationships between variable sets

region of rejection: the proportion of area in a sampling distribution that equals the level of significance. The region of rejection represents all those values of a test statistic that are highly unlikely provided that the null hypothesis is true

reliability: the property of a measure that consistently gives the same answer at different points in time

repeated measures design: a research design wherein measurements are made on the same subjects

replication: an independent verification of a research study

research question: a tentative generalization concerning the relationship between two or more variables

rough commercial: a "first draft" of a television or radio commercial that uses amateur actors, low-cost equipment, and rough editing

rough cut: a model or simulation of a final product

sample: a subgroup or subset of a population or universe

sample distribution: the frequency distribution of all the variables of interest as determined from a sample

sample-specific results: research results that are based on, or specific to, the research sample used

sampling distribution: a probability distribution of all possible values of a statistic which would occur if all possible samples of a fixed size from a given population were taken

sampling error: the degree to which measurements obtained from a sample differ from the measurements that would be obtained from the population

sampling frame: a list of the members of a particular population

sampling interval: a random interval used for selection of subjects or units in the systematic sampling method

sampling rate: the ratio of the number of people chosen in the sample to the total number in the population. For example, if 100 fraternity members were systematically chosen from a sampling frame of 1000 fraternity members, the sampling rate would be 10 percent or 1/10

scalar: a single digit (used in matrix algebra)

scattergram: a graphic technique for portraying the relationship between two variables

scientific research: a systematic, controlled, empirical, and critical investigation of hypothetical propositions about the presumed relations among natural phenomena

secondary analysis: a research method whereby researchers use data collected by another researcher or research organization (also called *data reanalysis*)

semantic differential: a rating scale consisting of seven spaces between two bipolar adjectives (for example, "good___:___:___:___:___:___:___bad")

share: an estimate of the percentage of persons or households tuned to a specific station or network

sigma (Σ): a Greek letter used as a symbol for "the sum of"

skewness: the degree of departure of a curve from the normal distribution (curves can be positively or negatively skewed)

sleeper effect: a situation where an attitude change that is minimal or nonexistent (and therefore unnoticed) in the short run may become significant in the long run

SMOG Grading: a measure of readability based on the number of syllables per word

social audit: in public relations research, an analysis of social performance of an organization

square matrix: a matrix in which the number of rows equals the number of columns

standard deviation: the square root of the variance (a mathematical index of dispersion)

standard error: an estimate of the amount of error present in a measurement

standard score: a measure that has been standardized in relation to a distribution's mean and standard deviation

storyboard: a device used to display photographs or artists' conceptions of the key scenes in a planned commercial

stratified sample: a sample selected after the population has been divided into categories

structured interview: an interview in which standardized questions are asked in a predetermined order

summary statistics: statistics that summarize a great deal of numerical information about a distribution such as the mean and standard deviation

sweep: a nationwide survey conducted by the A. C. Nielsen Company in which every television market is included

systematic variance: a systematic increase or decrease of all scores or data in a research study by a known factor

telephone coincidentals: a broadcasting research procedure whereby random subjects or households are called and asked questions about their present viewing or listening habits

telephone survey: a research method whereby survey data are collected over the telephone by trained interviewers who ask questions and record responses

theory: a set of related propositions that presents a systematic view of phenomena by specifying relationships among concepts

time series design: an analysis whereby subjects are tested or interviewed several times over a number of intervals (usually several weeks or months)

total observation: in field observation, a situation in which the observer assumes no role in the phenomenon being observed other than that of a pure observer

total participation: in field observation, a situation in which the observer becomes a full-fledged participant in the situation under observation

total survey area (TSA): a region where an audience survey is conducted

trend analysis: a longitudinal study in which the same topic is restudied but different groups of respondents are used, such as the Roper studies on the credibility of the media

tracking study: a special readership measurement technique whereby respondents are asked to designate by means of pencil markings the articles that they have read (using a different color of pencil for each reading episode)

t-test: a statistic used to determine significance between group means

Type I error: a rejection of the null hypothesis when it should be accepted

Type II error: an acceptance of the null hypothesis when it should be rejected

unit of analysis: the smallest element of a content analysis; the thing that is counted whenever it is encountered

unobtrusive research: any research conducted in such a way that the subjects are unaware of being observed

unstructured interview: an interview in which the interviewer asks broad and general questions but retains control over the discussion

uses and gratifications study: a study of the motives for media usage and the rewards that are sought

validity: the property of a test that actually measures what it purports to measure

variable: a phenomenon or event that can be measured or manipulated

variance: a mathematical index of the degree to which scores deviate from the mean

vector: a series of data points represented by a line in space

volunteer sample: a group of people who go out of their way to participate in a survey or experiment, for example, through a newspaper advertisement

weighting: a mathematical procedure used to adjust the sample to meet the characteristics of a given population

Name Index

Subject Index