Within this “Company-Customer-Competition” environment, many types of marketing research can be conducted, much of which is focused on using surveys for

- Monitoring customers and markets
- Measuring awareness, attitudes, and image
- Tracking product usage behavior
- Diagnosing immediate business problems
- Supporting strategy development

More specific examples are found in the Qualtrics.com Survey University. This provider of professional survey software identifies twenty different kinds of surveys that are of use to marketing researchers. Each focuses on a different aspect of the “Company” and it’s interaction with the “Customer” and “Competition” in the market environment:

Exhibit 1.1 Twenty Different Types of Marketing Surveys

1 - Market Description Surveys
To determine the size and relative market share of the market. Such studies provide key information about market growth, competitive positioning and tracking share of market.

2 - Market Profiling-Segmentation Surveys
To identify who the customers are, who they are not, and why they are or are not your customers. This is often a descriptive market segmentation and market share analysis

3 - Stage in the Purchase Process / Tracking Surveys
Where is the customer in the adoption process? This information shows market Awareness – Knowledge – Intention – Trial – Purchase – Repurchase of the product.

4 - Customer Intention - Purchase Analysis Surveys
Directed at understanding the current customer. What motivates the customer to move from interest in the product to actual purchase? This is a key to understanding customer conversion, commitment and loyalty.

5 - Customer Attitudes and Expectations Surveys
Does the product meet customer expectations? What attitudes have customers formed about the product and/or company. Used to direct advertising and improve customer conversion, commitment and loyalty.

6 - Customer Trust - Loyalty – Retention Analysis Surveys
Especially for high priced consumer goods with long decision and purchase processes (time from need recognition to purchase), and depth of consumer attitudes formed about the product and/or company.

7 - New Product Concept Analysis Surveys
Concept test studies are appropriate in the initial screening of new product concepts. Likes and dislikes about the concept and evaluation of acceptability and likelihood of purchase are especially useful measures.

8 - New Product Acceptance and Demand Surveys (Conjoint Analysis)
Primarily for estimating demand for new products that can be described or have been developed in drawing or concept, but have not yet been developed physically. Develops develop market share estimates of market potential for the alternative potential products.

9 - Habits and Uses Surveys
Directed at understanding usage situations, including how, when and where the product is used. Habits and uses studies sometimes include a real or virtual pantry audit.

10 - Product Fulfillment Surveys (Attribute, Features, Promised Benefits)
Evaluation of the product’s promised bundle of benefits (both tangible and image). Are expectations created for the product by advertising, packaging and the produce appearance fulfilled by the product?

11 - Product Positioning Surveys (Competitive Market Position)
A “Best Practices” study of “How does the market view us relative to the competition?” Competitive positioning analyses often compare the attributes and benefits that make up the product using multidimensional scaling.

2. The Environment or Context of the Problem

Consider the problem of deciding whether to introduce a new consumer product. The marketing researcher must work closely with the client in transforming the client's problem into a workable research problem.

The researcher's efforts should be oriented toward helping the manager decide whether any investigation is justified based on the potential value of the research findings versus their cost. The researcher must be aware of, and assist in, the identification of objectives, courses of action, and environmental variables, insofar as they affect the design of the research investigation.

If the research is undertaken and if the resulting findings are to be utilized (i.e., have an influence on the user's decision making), the manager and researcher must have a productive and trusting relationship that is based on the researcher's ability to perform and deliver the research as promised.

3. The Nature of the Problem

Every research problem may be evaluated on a scale that ranges from very simple to very complex. The degree of complexity depends on the number of variables that influence the problem. Understanding the nature of the problem helps a researcher ensure that the right problem is being investigated and that a marketing plan can be developed to solve the problem. A thorough preliminary investigation using focus groups of consumers, salespeople, managers, or others close to the problem may produce much needed insight.

4. Alternative Courses of Action

A course of action specifies a behavioral sequence that occurs over time, such as the adoption of a new package design, or the introduction of a new product. Such a program of action becomes a commitment, made in the present, to follow some behavioral pattern in the future.

It is usually desirable to generate as many alternatives as possible during the problem formulation stage and state them in the form of research hypotheses to be examined. A hypothesis often implies a possible course of action with a prediction of the outcome if that course of action is followed.

Once the nature of the problem has been agreed upon, the course of action must be specified. This involves:

1. Determining which variables affect the solution to the problem
2. Determining the degree to which each variable can be controlled
3. Determining the functional relationships between the variables and which variables are critical to the solution of the problem.

The following example shows the results of a failure to follow through with these aspects of the problem situation model.

EXHIBIT 1.3 "New Coke" Versus Original Coke

In the mid-1980s the Coca Cola Company made a decision to introduce a new beverage product (Hartley, 1995, pp. 129–145). The company had evidence that taste was the single most important cause of Coke's decline in the market share in the late 1970s and early 1980s. A new product dubbed "New Coke" was developed that was sweeter than the original-formula Coke.

Almost 200,000 blind product taste tests were conducted in the United States, and more than one-half of the participants favored New Coke over both the original formula and Pepsi. The new product was introduced and the original formula was withdrawn from the market. This turned out to be a big mistake! Eventually, the company reintroduced the original formula as Coke Classic and tried to market the two products. Ultimately, New Coke was withdrawn from the market.

What went wrong? Two things stand out. First, there was a flaw in the market research taste tests that were conducted: They assumed that taste was the deciding factor in consumer purchase behavior. Consumers were not told that only one product would be marketed. Thus, they were not asked whether they would give up the original formula for New Coke. Second, no one realized the symbolic value and emotional involvement people had with the original Coke. The bottom line on this is that relevant variables that would affect the problem solution were not included in the research.

CBS New Coke News Clip: http://www.youtube.com/watch?v=-doEpVWFLsE&NR=1&feature=fvwp
New Coke Commercial: http://www.youtube.com/watch?v=o4YvmN1hvNA
New Coke and Coke Classic Commercial: http://www.youtube.com/watch?v=ky45YGUA3co

5. The Consequences of Alternative Courses of Action

A set of consequences always relate to courses of action and even to the occurrence of events not under the control of the manager. One of the manager's primary jobs is to anticipate and communicate the possible outcomes of various courses of action that may result from following the research.

6. Degrees of Uncertainty

Most marketing problems are characterized by a situation of uncertainty as to which course of action is best. Years of experience may allow the decision-making manager to assign various "likelihoods of occurrence" to the various possible outcomes of specific courses of action.

A carefully formulated problem and statement of research purpose is necessary for competent research. The statement of purpose involves a translation of the decision maker's problem into a research problem and the derivation of a study design from this problem formulation. The research problem provides relevant information concerning recognized (or newly generated) alternative solutions to aid in this choice.

STAGE 2: METHOD OF INQUIRY

Market researchers look to the scientific method as the source of their investigative methods. Even though this method is not the only one used, it is the standard against which other investigative methods are measured. The scientific method makes great use of existing knowledge both as a starting point for investigation and as a check on the results of the investigations (i.e., a test of validity). Its most distinctive characteristic is its total lack of subjectivity. The scientific method has evolved objective and rigid procedures for verifying hypotheses or evaluating evidence. It is analytical in its processes and is investigator-independent. Thus, the scientific method is for the most part logical and objective, and frequently makes extensive use of mathematical reasoning and complicated experiments.

In addition, the cause and effect have to be related. That is, there must be logical implication (or theoretical justification) to imply the specific causal relation.

Associative Variation

Associative variation, or "concomitant variation," as it is often termed, is a measure of the extent to which occurrences of two variables are associated. Two types of associative variation may be distinguished:

1. Association by presence: A measure of the extent to which the presence of one variable is associated with the presence of the other

2. Association by change: A measure of the extent to which a change in the level of one variable is associated with a change in the level of the other.

It has been argued that two other conditions may also exist, particularly for continuous variables: (a) the presence of one variable is associated with a change in the level of other; and (b) a change in the level of one variable is associated with the presence of the other (Feldman, 1975).

Sequence of Events

A second characteristic of a causal relationship is the requirement that the causal factor occur first; the cause must precede the result. In order for salesperson retraining to result in increased sales, the retraining must have taken place prior to the sales increase.

Absence of Other Possible Causal Factors

A final basis for inferring causation is the absence of any other possible causes other than the one(s) being investigated. If it could be demonstrated, for example, that no other factors present could have caused the sales increase in the third quarter, we could then logically conclude that the salesperson training must have been responsible.

Obviously, in an after-the-fact examination of an uncontrolled result such as an increase in detergent sales, it is impossible to clearly rule out all causal factors other than salesperson retraining. One could never be completely sure that there were no competitor-, customer-, or company-initiated causal factors that would account for the sales increase.

Conclusions Concerning Types of Evidence

No single type of evidence, or even the combination of all three types considered, can ever conclusively demonstrate that a causal relationship exists. Other unknown factors may exist. However, we can obtain evidence that makes it highly reasonable to conclude that a particular relationship exists. Exhibit 2.1 shows certain questions that are necessary to answer.

To a large extent these major error components are inversely related. Increasing the sample size to reduce sampling error can increase non-sampling error in that, for example, there are more instances where such things as recording errors can occur, and the impact of biased (i.e., nonobjective) questions and other systematic errors will be greater. Thus, this inverse relationship lies at the heart of our concern for total error.

Ideally, efforts should be made to minimize each component. Considering time and cost limitations this can rarely be done. The researcher must make a decision that involves a tradeoff between sampling and non-sampling errors. Unfortunately, very little is known empirically about the relative size of the two error components, although there is some evidence that non-sampling error tends to be the larger of the two. In a study comparing several research designs and data collection methods, Assael and Keon (1982) concluded that non-sampling error far outweighs random sampling error in contributing to total survey error.

As an introduction, Exhibit 2.3 briefly defines eight major types of errors that can influence research results.

EXHIBIT 2.3 Types of Errors in the Research Process

Different types of errors can influence research results:

- **Population specification:** noncorrespondence of the required population to the population selected by the researcher
- **Sampling:** noncorrespondence of the sample selected by probability means and the representative sample sought by the researcher
- **Selection:** noncorrespondence of the sample selected by nonprobability means and the sought representative sample
- **Frame:** noncorrespondence of the sought sample to the required sample
- **Nonresponse:** noncorrespondence of the achieved (or obtained) sample to the selected sample
- **Surrogate information:** noncorrespondence of the information being sought by the researcher and that required to solve the problem
- **Measurement:** noncorrespondence of the information obtained by the measurement process and the information sought by the researcher
- **Experimental:** noncorrespondence of the true (or actual) impact of, and the impact attributed to, the independent variable(s)

Population Specification Error

This type of error occurs when the researcher selects an inappropriate population or universe from which to obtain data.

Examples: Cessna Aircraft conducts an online survey to learn what features should be added to a proposed corporate jet. They consider conducting a survey of purchasing agents from major corporations presently owning such aircraft. However, they believe that that this would be an inappropriate research universe; since pilots are most likely play a key role in the purchase decision.

Similarly, packaged goods manufacturers often conduct surveys of housewives, because they are easier to contact, and it is assumed they decide what is to be purchased and also do the actual purchasing. In this situation there often is population specification error. The husband may purchase a significant share of the packaged goods, and have significant direct and indirect influence over what is bought.

REFERENCES

Anderson, B. F. (1971). *The psychology experiment: An introduction to the scientific method* (2nd ed.). Belmont, CA: Brooks/Cole.

Assael, H., & Keon, J. (1982, Spring). Non-sampling vs. sampling errors in survey research. *Journal of Marketing, 46,* 114–123.

Feldman, J. (1975). Considerations in the use of causal-correlational technique in applied psychology. *Journal of Applied Psychology, 60,* 663–670.

Fern, E. F. (1982). The use of focus groups for idea generation: The effects of group size, acquaintanceship, and moderator on response quantity and quality. *Journal of Marketing Research, 19* (February), 1–13.

Fink, A. (2003). *How to design surveys* (2nd ed.). Thousand Oaks, CA: Sage.

Keeter, S., Miller, C., Kohut, A., Groves, R. M., & Presser, S. (2000). Consequences of reducing non-response in a national telephone survey. *Public Opinion Quarterly, 64,* 125–148.

Lilien, G., Brown, R., & Searls, K. (1991, January 7). How errors add up. *Marketing News, 25,* 20–21.

Monroe, K. B., & Petroshius, S. M. (n.d.). Developing causal priorities. Unpublished working paper, College of Business, Virginia Polytechnic Institute and State University.

Tomaskovic-Devey, D., Leiter, J., & Thompson, S. (1994). Organizational survey response. *Administrative Science Quarterly, 39,* 439–457.

Tull, D. S., & Albaum, G. S. (1973). *Survey research: A decisional approach*. New York: Intext Educational Publishers.

Yu, J., Albaum, G., & Swenson, M. (2003). Is a central tendency error inherent in the use of semantic differential scales in different cultures? *International Journal of Market Research, 45* (2), 213–228.

Internet Databases

The Internet has become the staple of research and provides access to most commercial electronic databases. Thousands of such databases are available from numerous subscription systems, such as DIALOG (http://www.dialog.com), LexisNexis (http://www.lexisnexis.com/), or Dow Jones News/Retrieval http://www.dowjones.com/Products_Services/ElectronicPublishing/EnterpriseMedia.htm.

In general, there are five categories of commercial databases:

1. Bibliographic databases that index publications
2. Financial databases with detailed information about companies
3. Statistical databases of demographic, econometric, and other numeric data for forecasting and doing projections
4. Directories and encyclopedias offering factual information about people, companies, and organizations
5. Full-text databases from which an entire document can be printed out.

The advantages of such current databases are obvious. All that is needed is personal computer with internet access or a CD-ROM/DVD.

Computerized databases have led to an expanded focus on database marketing. Database marketing has been defined as an extension and refinement of traditional direct marketing, but uses databases to target direct response advertising efforts and tracks response and/or transactions. In database marketing, the marketer identifies behavioral, demographic, psychographic, sociological, attitudinal, and other information on individual consumers/households that are already served or that are potential customers. Data may come from secondary and/or primary sources. Qualtrics clients are increasingly APIs (Application Programming Interface) to link and integrate customer databases with survey data and respondent panels. APIs can be used to link and integrate data from multiple sources in real time. Thus, information in database profiles is augmented by new contact and survey data, and can be viewed in dashboards that report current information and can be used to better target and predict market response. Databases can be used to estimate market size, find segments and niches for specialized offerings, and even view current customer use and spending (Morgan, 2001). In short, it helps the marketer develop more specific, effective, and efficient marketing programs.

Today, data mining is in high demand as a research focus. Data mining involves sifting through large volumes of data to discover patterns and relationships involving a company's products and customers. Viewed as a complement to more traditional statistical techniques of analysis, two of the more powerful data mining techniques are neural networks and decision trees (Garver, 2002). Further discussion of data mining techniques is beyond the scope of this text, but good discussions are found in Berry & Linoff (1997, 2002) and Dehmater & Hancock (2001).

> For example, ACNielsen uses a metering device that provides information on what TV shows are being watched, how many households are watching, and which family members are watching. The type of activity is recorded automatically; household members merely have to indicate their presence by pressing a button. The sample is about 5,000 households. In local markets, the sample may be 300 to 400 households.

Single source data tracks TV viewing and product purchase information from the same household. Mediamark's national survey and IRI's BehaviorScan are examples of such single-source data. The single-source concept was developed for manufacturers who wanted comprehensive information about brand sales and share, retail prices, consumer and trade promotion activity, TV viewing, and household purchases.

The information obtained from the types of syndicated services described previously has many applications. The changes in level of sales to consumers may be analyzed directly without the problem of determining changes in inventory levels in the distribution channel. Trends and shifts in market composition may be analyzed both by type of consumer and by geographic areas. A continuing analysis of brand position may be made for all brands of the product class. Analyses of trends of sales by package or container types may be made. The relative importance of types of retail outlets may be determined. Trends in competitor pricing and special promotions and their effects can be analyzed along with the effects of the manufacturer's own price and promotional changes. Heavy purchasers may be identified and their associated characteristics determined. Similarly, innovative buyers may be identified for new products and an analysis of their characteristics made to aid in the prediction of the growth of sales. Brand-switching and brand-loyalty studies may be made on a continuing basis. One reported use of this syndicated service has been to design products for specific segments.

The products from syndicated services are continually changing with client needs and new technological opportunities.

SUMMARY

This chapter has been concerned with secondary information and sources of such information. We started with some reasons why secondary information is essential to most marketing research projects. Then, various sources and types of secondary information—internal and external—were discussed in some depth. Also given more than cursory treatment was syndicated data, a major type of service provided by commercial agencies.

Simmons Study of Media and Markets. http://www.libs.uga.edu/ref/simmons/index.html
Forty volumes devoted to the determination of use of various products and services.

Standard Rate and Data Service. http://www.srds.com/portal/main?action=LinkHit&frameset=yes&link=ips
This service offers separate directories giving advertising rates, specifications, and circulation for publications, broadcast stations, and so forth in the following media: business publications (monthly), community publications (semiannual), consumer magazines and agri-media (monthly), newspapers (monthly), print media production (quarterly), spot radio (monthly), spot television (monthly), and weekly newspapers (monthly). This service also publishes an annual *Newspaper Circulation Analysis* covering newspaper circulation and metro area, TV market, and county penetration. Entries include daily and Sunday circulation figures, by county, for each newspaper.

Company and Industry Data

Dun and Bradstreet. *Million Dollar Directory*. Annual.
Lists over 160,000 American companies having an indicated net worth of $500,000 or more.

Standard and Poor's Register of Corporations, Directors and Executives. Annual.
Lists over 45,000 U.S. businesses.

Directory of Corporate Affiliations. (Annual) and *America's Corporate Families: The Billion Dollar Directory* (Annual).
Each directory features lists of divisions and subsidiaries of parent companies.

MOODY's Investor Service. Various call numbers. Seven annual volumes plus twice-a-week supplements. Includes Bank and Finance, Industrial, Municipal and Government, Over-the-Counter, Public Utility, Transportation, and International.

Standard and Poor's Corporation Records.

Value Line Investment Survey. Weekly.

Walker's Manual of Western Corporations and Services. Annual.

Standard and Poor's Corporation. *Industry Surveys*.
This loose-leaf service includes for 22 industry categories detailed analysis of each category and of the industries that comprise it.

Predicasts Forecasts. Quarterly.
Is useful both as a source of forecasts and projections of industries, products, and services and as a finding aid to other sources.

U.S. Industrial Outlook. Annual.
Contains short-term forecasts for specific industries.

International Business

Many sources listed previously also are relevant to foreign markets and international marketing activities.

Indexes and Guides

Predicasts Funk and Scott Indexes.

1. *F & S Europe*. Monthly.
Covers articles or data in articles on foreign companies, products and industries.

2. *F & S International Index*. Monthly.
Covers articles or data in articles on foreign companies, products, and industries.

4. **Disclosure.** The willingness on the part of both parties to reveal their “true” selves to one another.
5. **Feedback.** The continuous stream of verbal and nonverbal signals (e.g., smiles, puzzled expressions, raised eyebrows, moans) sent between interview parties that reveal feelings, belief or disbelief, approval or disapproval, understanding or misunderstanding, interest or disinterest, and awareness or
6. **Cooperation.** The degree to which the interview parties are willing and able to reduce the competition inherent in most interview situations and work together for their mutual benefit.
7. **Conflict.** The potential or actual struggle between parties because of incompatible or opposing needs, desires, demands, and perceptions.
8. **Trust.** Belief in the good, worth, ethics, believability, and reliability of the other party.

Involvement, Control and Relationships have some effect upon each of the elements. These dimensions and elements of relationships are present in each interview but are not of equal importance. Although they are independent of each other, they have strong interdependence as well.

SOURCE: From Stewart, C. J. and Cash, W. B., *Interviewing Principles and Practices*, 4th ed. © 1985 William C. Brown, Publishers. Reprinted with permission of The McGraw-Hill Companies, pp. 9–13

Structure of the Interview

Interviews in marketing research and the behavioral sciences typically involve information gathering and are usually classified by two major characteristics. An interview is either **structured or unstructured**, depending on whether a formal questionnaire has been formulated and the questions asked in a prearranged order. An interview is also categorized as either **direct or indirect**, reflecting whether the purposes of the questions are intentionally disguised. Cross-classifying these two characteristics helps us to identify four different types of interviews

Objective Interviews (a) structured and direct
(b) unstructured and direct

Subjective Interviews (c) structured and indirect
(d) unstructured and indirect.

Types a and b are basically objectivist; types c and d, subjectivist. We discuss each type of interview in turn (although the discussion of the two indirect types of interviews is combined). We then discuss the media through which interviews may be conducted.

Structured-Direct Interviews

Structured-direct interviews are the usual type of consumer survey to “Get the Facts” and obtain descriptive information. A formal questionnaire is used consisting of nondisguised questions.

> **Example**: A marketing research manager of a bedroom-furniture manufacturer wants to find out how many and what kinds of people prefer various styles of headboards and dressers. The question sequence is fixed only those questions are asked. The resulting interview is structured-direct in nature.

To use the bedroom furniture example again, if the owner of a bedroom set is asked the free-answer question, “Why did you buy your bedroom set?” the answer is almost certain to be incomplete, provide proximate causes and may be worthless.

If the interviewer were seeking motivations, consider answers such as “Because we needed a bed,” “Our old bed was worn out,” or “Because it was on sale.” When motivations are given, such as “We enjoy a comfortable mattress that gives us a good night’s sleep,” they are rarely complete.

The added enjoyment may be because the mattress is firmer, because of the pillow top, because of the prestige the owner attaches to having a carved oak bedroom set, or some combination of these and other factors. In addition, it is probable that motives other than “enjoyment” influenced the purchase.

When used to establish motives, the unstructured-direct interview is known as a ***depth interview***. The interviewer will continue to ask probing questions: “What did you mean by that statement?” “Why do you feel this way?” “What other reasons do you have?” The interviewer continues with similar questions until satisfied that all the information that can be obtained has been obtained, considering time limitations, problem requirements, and the willingness and ability of the respondents to verbalize motives.

The unstructured interview is free of the restrictions imposed by a formal list of questions. The interview may be conducted in a seemingly casual, informal manner in which the flow of the conversation determines which questions are asked and the order in which they are raised. The level of vocabulary used can be adapted to that of the respondent to ensure that questions are fully understood and rapport is developed and maintained. The flexibility inherent in this type of interview, when coupled with the greater informality that results when it is skillfully used, often results in the disclosure of information that would not be obtained in a structured-direct interview.

In the unstructured interview, the interviewer must both formulate and ask questions. The unstructured interview can therefore be only as effective in obtaining complete, objective, and unbiased information as the interviewer is skilled in formulating and asking questions. Accordingly, the major problem in unstructured direct interviews is ensuring that competent interviewers are used. Higher per-interview costs result, both as a result of this requirement and the fact that unstructured interviews generally are longer than those that use a questionnaire. In addition, editing and tabulating problems are more complicated as a result of the varied order of asking questions and recording answers.

Structured-Indirect and Unstructured-Indirect Interviews

A number of techniques have been devised to obtain information from respondents by *indirect* means. Both structured and unstructured approaches can be used. Many of these techniques employ the principle of *projection*, in which a respondent is given a non-personal, ambiguous situation and asked to describe it. It is assumed that the respondent will tend to interpret the situation in terms of his or her own needs, motives, and values. The description, therefore, involves a projection of personality characteristics to the situation described. These techniques are discussed in more depth in Chapter 5.

the problem. In addition, control procedures can and should be established to reduce it even more.

The simplest control procedure is to *call-back* a subsample of respondents. If the information on an initial interview is found to disagree significantly with that on the call-back interview, additional call-backs may be made on respondents originally interviewed by the same person. The fear of being caught will discourage cheating.

Other control procedures include the use of "cheater" questions and the analysis of response patterns. Cheater questions are informational questions that will disclose fabricated answers with a reasonably high probability of success. Likewise, the analysis of patterns of responses for interviewer differences will disclose interviewer cheating when significant variations from expected norms occur. Such analyses can be made at very little additional cost.

Time Costs

Perhaps the most common reason for respondent unwillingness to provide accurate information, or any information for that matter, is the result of the time required to make the information available. Respondents often give hasty, ill-considered, or incomplete answers and resist probing for more accurate information. When possible to do so, a respondent will tend to act in a manner that will reduce time costs. Such behavior often results in inaccurate or missing information.

When conducting telephone and personal interviews the interviewer might ask "Is this a good time to answer some questions, or would you rather set a time when I could contact you again?" Experience has shown this latter technique only slightly lowers response rates.

Perceived Losses of Prestige

When information attributing prestige to the respondent is sought, there is always a tendency to receive higher-prestige responses. All researchers experience difficulty both in recognizing the items that demand prestige content, and in measuring the resulting amount of inaccuracy. Information that affects prestige is often sensitive information, including socioeconomic (age, income, educational level, and occupation), place of birth or residence.

> An example of a still more subtle prestige association occurred in a study on nationally known brands of beer. One of the questions asked was, "Do you prefer light or regular beer?" The response was overwhelmingly in favor of light beer. Since sales data indicated a strong preference for regular beer, it was evident that the information was inaccurate. Subsequent investigation revealed that the respondents viewed people who drank light beer as being more discriminating in taste. They had, therefore, given answers that, in their view, were associated with a higher level of prestige.

Measuring the amount of inaccuracy is a difficult task. One solution to this problem is to ask for the information in two different ways. For example, when obtaining information on respondents' ages, it is a common practice to ask early in the interview, "What is your present age?" and later "In what year did you graduate high school?"

In one study, when respondents were asked, "Are you afraid to fly?" Very few people indicated any fear of flying. In a follow-up study, when they were asked, "Do you think your neighbor is afraid to fly?" (a technique known as the *third-person technique*), most of the neighbors turned out to have severe anxieties about flying.

__Invasion of Privacy__

Clearly, some topics on which information is sought are considered to be private matters. When such is the case, both nonresponse and inaccuracy in the responses obtained can be anticipated. Matters about which respondents resent questions include money matters or finance, family, life, personal hygiene, political beliefs, religious beliefs, and even job or occupation. It should be recognized however, that invasion of privacy is an individual matter. Thus, information that one person considers sensitive may not be viewed that way by others.

The investigator should attempt to determine sensitivity if it is suspected to be a problem. One way of handling this is to add questions in the pretest stage which ask about the extent of sensitivity to topics and specific questions. A comprehensive treatment of sensitive information and how to ask questions about it is given by Bradburn and Sudman (1979).

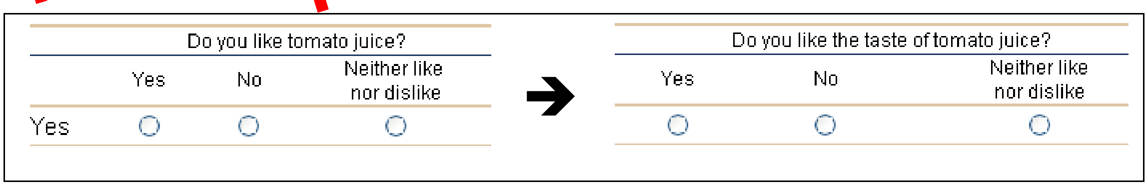
Ambiguity

Ambiguity includes errors made in interpreting spoken or written words or behavior. Ambiguity, therefore, occurs in the transmission of information, through either communication or observation.

__Ambiguity in Communication__

Ambiguity is present in all languages. Unambiguous communication in research requires that the question asked and the answers given each mean the same thing to the questioner and the respondent.

The first step in this process is the controlling one. If the question is not clearly understood by the respondent, frequently the answer will not be clearly understood by the questioner. To illustrate this point, after pretesting in an actual research project on tomato juice, the following question change occurred after pretesting.



Even a careful reading of these two questions may not disclose any real difference in their meaning. The analyst who drew up the question assumed that “like” refers to taste. In pretesting, however, it was discovered that some housewives answered “Yes” with other referent in mind. They “like” the amount of Vitamin C their children get when they drink tomato juice, they “liked” the tenderizing effect that tomato juice has when used in cooking of meat dishes, and so on. If the wording of the question had not been changed, there would have been a complete misunderstanding in some cases of the simple, one-word answer “Yes.”

A related issue is one where a shortened form of a sentence is used. Examples are, “How come?”, “What?”, and “How?” This “elliptical sentence” requires the respondent to first consider the context of the sentence and then add the missing parts. When the mental process of transformation is different for the researcher and respondent, communication is lost and interpretation of a person’s response is faulty and ambiguous.

Electronics driven lifestyles that include online social networks, massive use of texting and pervasive internet connections are no doubt responsible in part for seemingly responsible for attitude and behavioral changes in the way we view our increasingly virtual world. Strong upward trends are observed in the percentage of Internet purchases for airline tickets, CDs, DVDs, books, computer software, hardware and systems. These online customers provide excellent access for research purposes.

Advocates of online surveying quickly point to the elimination of mailing and interviewing costs, elimination of data transcription costs, and reduced turnaround time as the answer to client demand for lower cost, timelier, and more efficient surveys. As a result, online marketing research has become so widely accepted that online research has been optimistically projected to account for as much as half of all marketing research revenue, topping $3 billion. While these numbers appear to be overly optimistic, it is clear that online research is growing and that researchers operate in a much faster-paced environment than ever before. This pace will continue to increase as new modalities for research open: wireless PDAs, Internet-capable mobile phones, Internet TVs, and other Internet-based appliances yet to be announced. Each is an acceptable venue for interacting with the marketplace and conducting online research.

Substantial benefits accrue from the various approaches to computer-assisted data collection in surveys as shown below:

- Respondents need few computer related skills
- Respondent choose their own schedule for completing survey
- Can easily incorporate complex branching into survey
- Can easily pipe and use respondent generated words in questions throughout the survey
- Can accurately measure response times of respondents to key questions
- Can easily display a variety of graphics and directly relate them to questions
- Eliminates need to encode data from paper surveys
- Errors in data less likely, compared to equivalent manual method
- Speedier data collection and encoding compared to equivalent manual method.

ACNielsen (Miller, 2001) reported the results of 75 parallel tests comparing online and traditional mall intercept methods. Researchers noted high correlations in aggregate purchase intentions. While online measures may yield somewhat lower score values, recalibration of averages against appropriate norms produced accurate sales forecasts. Wilkie further reported that while responses gathered using different survey modes may be similar, the demographic profiles of online and traditional respondents groups do differ.

Given that the current percentage of households online is approximately 75 percent, statistical weighting of cases could be used to adjust demographic differences of online groups to match mall intercept or telephone populations. However, the possibility of weighting actually raises the question of whether it is better to model phone or mall intercept behavior (which are also inaccurate) or to attempt to independently model the actual behavior of the respondents.

was identified in the groups, the questioning centered on isolating the elements that made people change their minds.

Example: is the Jacksonville, Florida, symphony orchestra's use of focus groups to identify lifestyle marketing issues to explore entertainment alternatives, and to provide some ideas about what future audiences would want and expect from the orchestra (LaFlamme, 1988).

Raymond Johnson (1988) has identified four distinctive categories of focus groups on the basis of examining tapes from the project files of several research companies. Johnson, a practitioner, has defined each type of focus group by the adaptation of an interviewing technique to answer one of four basic research questions. The focus group types are as follows:

- *Exploratory studies* of consumer lifestyles and probing to "just find out what's on the consumers' minds these days."

- *Concept testing studies* of how a group, without prompting, interprets a deliberately sketchy idea for a new product or service. Potential users are able to react to a concept still in its formative or experimental stage.

- *Habits and usage studies* deal with the real world of actual consumers. The topic is framed by the moderator's instructions to describe, usually by situation-specific narratives, the details of personal experiences in using a particular product or service.

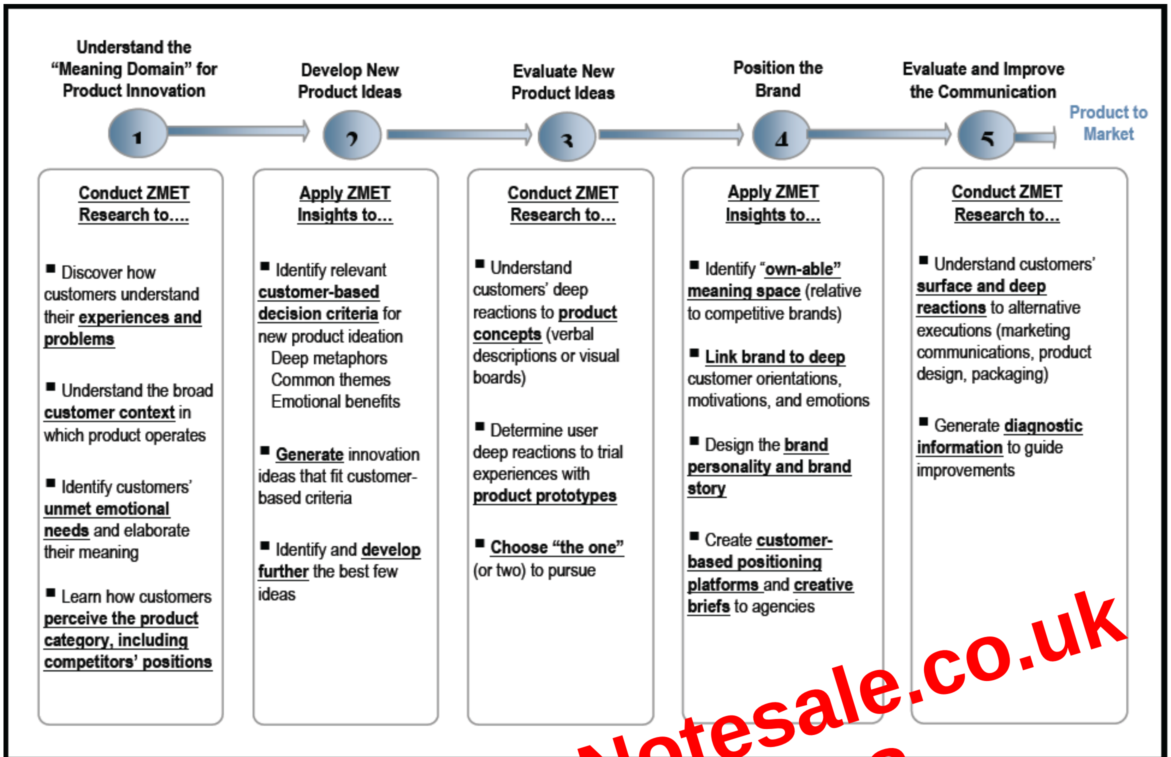
- *Media testing* in which participants are asked to interpret the message covered in media usually seen in rough form is the fourth type. All types of media may be covered. Group members talk about their understanding of the message and evaluate the extent to which they find it credible, interesting, and emotionally involving.

The use of focus groups is not limited to consumer products and services. This technique can provide a relatively easy and cost-effective way to interact with business consumers in industries ranging from pharmaceuticals to computer software. The ways in which focus groups are structured and conducted are similar for consumer-based and business-to-business groups, except as identified by Fedder (1990).

A natural question, of course, is, "Why do focus groups work?" One view is that clients are provided with a gut-level grasp of their customers. This means that a sense of what is unique about customers is gained —their self-perceptions, desires, and needs that affect everything they do. For more detailed discussions, see Bloor, Frankland, Thomas, and Robson (2001), and Fern (1981). Qualitative research offers not just an intellectual comprehension of consumers but a vivid, visceral recognition that affects, on a very deep level, how clients see, feel about, and deal with their customers from then on. Some guidelines and questions to assist clients in observing focus groups more effectively are discussed briefly in Exhibit 6.1.

One critical aspect of a focus group's success is the moderator (Exhibit 6.2). The moderator's job is to focus the group's discussion on the topics of interest. To accomplish this, the moderator needs to explain to the group the operational procedures to be used, the group's purpose, and why the group is being recorded. Rapport with the group must be established and each topic introduced. The respondents need to feel relaxed early on, and often moderator humor helps this to happen. Once the respondents are comfortable, the moderator needs to keep the discussion on track while not influencing how the discussion proceeds. A moderator has

Figure 6.1 Applications of ZMET in Product Marketing



Source: Olson, Waltersdorff and Forr (2008)

Table 6.2 shows examples of these deep metaphors that each of us in expressing thoughts and emotions. Interviewees are asked to explain their true feelings about objects of concern with images and not just words.

Table 6.2 Deep Metaphors and Expressions in Conversation

DEEP METAPHORS	EXPRESSIONS IN THE VIEW
Physicality	References to bodily functions and senses such as taste it; feel it; pick up, ingest, see my point; hurts me
Orientation	References to spatial orientation--up/down, higher/lower; bigger/smaller; upright/lie down; front/back.
Pleasure/pain	References to the positive (or negative); to things that give one pleasure; feeling good or bad; References to enjoyment, fun, happiness, euphoria, well-being ... or to he opposite (fear, disgust, pain); Feeling good versus hurting, physically or emotionally.
Entity	Considering an intangible idea, concept as a thing, a physical object, an entity; e.g.; "I can't get my breath"
Balance	References to equilibrium, stability, equalize or compensate; Including both sides; Images of scales, teeter-totter, balance beam; References to reciprocity--give and take; References to 'stable" emotional states such as calm, relaxed, serene; Feeling 'right" with the world.
Connection or Linkage	References to connecting to things or people; Making an association; References to linking or attaching; To be a part of; to not be isolated from; Liking or loving someone or something; References to getting in touch with yourself; find your true self.
Resource	References to having/getting the requisite knowledge, energy, tools, or materials to accomplish some task; Having or getting help and assistance from others (we are a team, I need support);

Container	References to being in (or out) of a place (house, room); References to keeping or storing; References to "in" and "out;" Keeping things out as well as in; Being wrapped up ... or out in open.
Motion or Movement	References to moving (flowing, traveling, running or walking); References to movement actions (hurrying, getting going); Keep moving; Keep it going.
Journey	References to taking a trip; Following a path, choosing a direction; Getting there; Journey of life.
Transformation	References to changing from one state to another-physical or emotional; Becoming something or someone else; References to evolving, maturing, growing;
Time	References to the passage of time; Reference to past events or historical perspectives; Images of clocks and watches; References to old memories, remembering past events.
Nature	References to nature, outdoors, natural world; Specific images of nature--rain forest, desert, woods; References to pure, unadulterated, pristine, uncontaminated and to wild, untamed, chaotic, stormy. References to breeding, evolving, growing.
Force	References to power, a powerful presence, or a source of energy; References to the consequences of force (getting hit; slammed, impact)
Fight vs. Flight	References to war; fights, battle, attacking, aggression; Choose your battles; References to weapons; Avoid a fight; Don't get involved; Running away or hiding from something; Ignoring an issue.
Knowing and knowledge	References to knowing, understanding, learning. Gaining knowledge and understanding through study; Knowledge and wisdom; Comprehension; insight. In contrast, the unknown, ignorance, stupidity; Inability to comprehend and understand something.
Contrast or Paradox	References to looking at both/two sides; Comparing opposites; Juxtaposition of opposites; References to paradox--being one thing and its opposite at same time;
Personal expression	Reference to things that express one's personal goals, values, points of view ... to one's self or to others.
The Ideal	Reference to the ideal object, situation, feeling. Statements about one's ideal self. References to perfection, the perfect one.
System	References to machine metaphors (wheels and gears, well-oiled machine); or a constructed process or approach for solving a problem; A set of rules or procedures to accomplish a task; Following a ritual
Sacred (Profane)	References to divine or spiritual qualities; Symbolic cleanliness, purity, vs. dirty, impure, devilish, malevolent
The Mother; Care giving	References to love, fondness; warm relationships with family and friends; Caring for or nurturing others, animals, plants; Being a caring resource for others; taking responsibility for and supporting and helping others
Masculine/ Feminine	References to gender--male or female; Having masculine or feminine traits; Attributions of gender to nonliving things, ideas, concepts.
Birth/Death	References to beginning or end of life, of an idea, or concept; Being rejuvenated or brought back to life (reborn); References to decline, dying, or death.
Quantify/ Measure	References to quantification; Keeping track; Counting; Yardsticks, scales, sextons; Computing amounts and quantities
Holism/ Completeness	References to being whole, entire, complete, not lacking anything, not having weaknesses that compromise any other part of the whole.

Source: Jerry C. Olson, Olson Zaltman Associates (www.olsonzaltman.com).

As a final note for conducting in-depth interviews, it is important to use empathy in understanding and appreciating someone else's beliefs, attitudes, feelings, and behaviors, as well as the social factors that influence their behavior. Many standard qualitative techniques neglect empathy. Specific guidelines for conducting empathic interviews include the following (Lawless, 1999):

Exhibit 6.4 discusses the use of *protocols* (Ericsson & Simon, 1984). This technique allows respondents to respond freely without intervention of an interviewer.

EXHIBIT 6.4 Protocols for the Qualitative Research Tool Kit

A protocol is a record of a respondent's verbalized thought processes while performing a decision task or while problem solving. This record is obtained by asking the respondent to "think out loud" or talk about anything going through his or her head while performing the task. Protocols can be collected either in a laboratory situation while the respondent is making a simulated purchase or in the field while an actual purchase decision is being made.

Protocols can be recorded concurrent with the task or retrospective when the verbalizing aloud is done just after the task has been finished.

In contrast to traditional survey methods, protocol methodology allows a person to respond freely in his or her own terms in relation to the actual choice task or decision situation. The form and particular stimuli to which the research subject should respond is not defined or specified by the researcher.

Protocols can be useful in studying brand choice, product categorization, product usage patterns and attitudes, and the impact of shopping environment and situational variables on behavior.

Example: Comcast, a major provider of cable TV, telephone and Internet services offered an unsuccessful self install kit that often resulted in customers calling for a technician to complete the installation. A protocol analysis was conducted when individuals in a laboratory situation were give a self-install kit and asked to complete an install while narrating their activities. Comcast identified the key problem areas in the install process and was able to revise installation instructions and provide an 800 support line, thereby making significant reduction in install service calls.

Story completion, an extension of the sentence completion technique, consists of presenting the beginning of a situational narrative to a respondent, who is asked to complete it. The general underlying principle is that the person will project his or her own psychological interpretation of the situation into the response. For example, the situation could be formulated as follows: "Last weekend my partner and I were deciding which jewelry store to visit for a purchase. When I mentioned XYZ, my partner remembered the last visit there. Now you complete the story."

OBSERVATION

The remaining major method of collecting qualitative information is through observation. Observation is used to obtain information on both current and past behavior of people. Rather than asking respondents about their current behavior, it is often less costly and/or more accurate if the behavior is observed. We clearly cannot observe past behavior, but the results of such behavior are often observable through an approach known as the case study or customer case research. This exploratory qualitative methodology traces the stories of people, circumstances, decisions and events leading to actual purchase decisions through one-on-one interviews (Berstell & Nitterhouse, 2001). The case study approach allows for determining any underlying patterns and may uncover unforeseen problems and unexpected opportunities. Some key characteristics of this approach are (Berstell, 1992):

- Case studies uncover motivations through demonstrated actions, not through opinions.
- Studies are conducted where a product is bought or used.

We will discuss each of these areas.

Applicability of Indirect Research Techniques

The basic premises leading to the use of indirect research techniques are as follows:

1. The criteria employed and the evaluations made in most buying and use decisions have subconscious thoughts and emotions.

2. This subconscious content is an important determinant of what we hear, feel, think, say and do when performing a choice behavior such as buying a product.

3. Such content is not adequately or accurately verbalized by the respondent through direct communicative techniques.

4. Such content is adequately and accurately verbalized by the respondent through indirect communicative techniques.

How valid are these premises? We have already seen that they are valid for some problems. While correct, this is not a very satisfactory answer. It is more useful to review situations in which indirect research information might reasonably be sought from respondents. Four situational categories can be distinguished in which information might be sought from respondents.

Category one is where the information desired is known to the respondent and he or she will give it if asked. Direct questioning will therefore provide all of the needed information in this situation. If the reason a consumer does not buy brand X tires is because he believes they do not wear as well as they should, he will willingly say so given the opportunity.

Category two is where the information desired is known to the respondent, but he or she does not want to divulge it. Matters that are considered to be private in nature, that are believed to be prestige- or status-bearing, or that are perceived as presenting a potential respondent-investigator opinion conflict may not be answered accurately. That otherwise quiet and retiring people sometimes buy powerful cars because it gives them a feeling of superiority on the highway are not reasons that will likely be expressed openly and candidly in response to direct questions. When underlying motivations of this general nature are believed to exist, indirect techniques are well suited to elicit such information.

Third, the information desired is obtainable from the respondent, but he or she is unable to verbalize it directly. When respondents have reasons they are unaware of, such as the association of the use of instant coffee with lack of planning and spendthrift purchasing, properly designed and administered indirect techniques can be highly useful for uncovering such motivations.

Fourth, the information desired is obtainable from the respondent only through inference from observation. In some cases motivations of respondents are so deep-seated that neither direct nor indirect methods of questioning will bring them to the surface.

> An experiment in which the same detergent in three different-colored boxes resulted in the opinion of housewives using them that the detergent in the blue box left clothes dingy, that the one in the yellow box was too harsh, and that the one in the blue-and-yellow box was both gentle and effective in cleaning is an illustration of color association and its effect on assessment of

Census, or Sample?

Once the population has been defined, the investigator must decide whether to conduct the survey among all members of the population, or only a sample subset of the population. The desirability and advantages of using a sample rather than a census depend on a variety of factors such as geographic location, the absolute size of the population, and the sample size required for results sufficiently accurate and precise to achieve the required purposes.

Two major advantages of using a sample rather than a census are speed and timeliness. A survey based on a sample takes much less time to complete than one based on a census. Frequently, the use of a sample results is a notable economy of time, money and effort, especially when a census requires hiring, training and supervising many people.

In other situations, a sample is necessary because of the destructive nature of the measurement, such as in product testing. There is a related problem over surveying human populations when many different surveys need to be conducted on the same population within a relatively short period of time. Nonprobability sampling techniques explicitly protect against this problem.

In still other situations, a sample may control non-sampling errors. Samples (a smaller number of interviews compared to a census) may result in better interviewing, higher response rates through more call backs, and better measurement in general. The total amount of sampling and non-sampling error of a sample may actually be than the non-sampling error alone would be for a census. In Chapter 2, we emphasized the importance of minimizing total error.

However, under certain conditions a census may be preferable to a sample. When the population is small, the variance in the characteristic being measured is high, the cost of error is high, or the fixed costs of sampling are high, sampling may not be useful. In addition, if the characteristic or attribute of interest occurs rarely in the population, then a census of a tightly defined population might be desirable (for example, people with a rare genetic disorder). In this case, it would be necessary to sample a relatively large proportion of the general population to provide statistically reliable information. Obviously, the practicality of this depends upon the absolute size of the population and the occurrence rate for the characteristic of interest.

Sample Design

Sample design is "the theoretical basis and the practical means by which data are collected so that the characteristics of a population can be inferred with known estimates of error" (Survey Sampling, Inc.).

Operationally, sample design is at the heart of sample planning. Sample design specification, including the method of selecting individual sample members, involves both theoretical and practical considerations (such as cost, time, labor involved, organization). The following checklist is suggested to obtain a sample that represents the target population (Fink, 2003):

1. Are the survey objectives stated precisely?
2. Are the eligibility criteria for survey respondents or experimental subjects clear and definite? Exclusion criteria rule out certain people.
3. Are rigorous sampling methods chosen? This involves selecting an appropriate probability or nonprobability sampling method.
4. Further questions to be answered in this section include:

- What type of sample should be used?
- What is the appropriate sampling unit?
- What is the appropriate frame (that is, list of sampling units from which the sample is to be drawn) for the particular design and unit decided upon?
- How are refusals and nonresponse to be handled?

Type of Sample

Much of the sampling in the behavioral sciences and in marketing research, by its nature, involves samples are selected by the judgment of the investigator, convenience, or other nonprobabilistic (nonrandom) processes. In contrast, probability samples offer the promise of bias-free selection of sample units and permits the measurement of sampling error. Nonprobability samples offer neither of these features. In nonprobability sampling one must rely on the expertise of the person taking the sample has selected respondents in an unbiased manner, whereas in probability sampling the sampling of respondents is independent of the investigator.

> **Example:** A dog food manufacturer tested consumer reactions to a new dog food by giving product samples to employees who own dogs and eliciting their responses about a week later. The employees' dogs liked the food and the pet food manufacturer actually introduced the new dog food product. However when it hit the market the product was a flop… dogs simply would not eat it. As managers scrambled to find out what went wrong, research showed that employees were so loyal to the company's products that their dogs had little variety and would eat anything for a change. In the broader market dogs were used to a greater variety of dog foods including table scraps and just did not like the taste of the new dog food. In this case, a biased sample was erroneously assumed to conform to the general population of dogs and dog owners.

A researcher choosing between probability and nonprobability sampling implicitly chooses the probability sample's relative size of sampling error against the nonprobability sample's combined sampling error and selection bias. For a given cost, one can usually select a larger nonprobability sample than probability sample, meaning that the sampling error should be lower in the nonprobability sample, but that the nonrandom process used for selecting the sample may have introduced a selection bias.

The Sampling Unit

The sampling unit is the unit of the population to actually be chosen during the sampling process. The sampling unit may contain one or more elements describing the population. For instance, a group medical practice may be interested in surveying past patient behavior of the male wage earner *or* his entire household. In either case, it may be preferable to select a sample of households as sampling units.

The Sampling Frame

A sampling frame is a means of identifying, assessing and selecting the elements in the population. The sampling frame usually is a physical listing of the population elements. In those instances where such a listing is not available, the frame is a procedure producing a result equivalent to a physical listing. For instance, in a consumer survey where personal interviews are conducted with a mall intercept method is used to obtain data, the sample frame includes all those people who enter the mall during the study period.

Nevertheless, this method continues to be widely used due to the segmentation of markets that companies routinely engage in.

The Cluster Sample

The researcher will ordinarily be interested in the characteristics of some elementary element in the population such as individual family attitudes toward a new product. However, when larger primary sampling units are desired, cluster sampling may be used. For example, the researcher may choose to sample city blocks and interview all the individual families residing therein. The blocks, not the individual families, would be selected at random. Each block consists of a cluster of respondents. The main advantage of a cluster sample relative to simple random sampling is in lower interviewing costs rather than in greater reliability.

The Area Sample: Single Stage and Multistage

As the name suggests, area sampling pertains to primary sampling of geographical areas—for example, counties, townships, blocks and other area descriptions. A single-stage area sample occurs when only one level of sampling takes place (such as a sampling of blocks) before the basic elements are sampled (the households). If a hierarchy of samples within the larger area is taken before settling on the final clusters, the resulting design is usually referred to as a multistage area sample.

> Example: Consider the sample design used by the Gallup Organization for taking a nationwide poll. Gallup draws a random sample of locations at the first stage of the sampling process. Blocks or geographic segments are then randomly sampled from each of these locations in a second stage, followed by a systematic sampling of households within the blocks or segments. A total of about 1,500 persons are usually interviewed in the typical Gallup poll.

METHODS FOR DETERMINING SAMPLE SIZE

There are several ways to classify techniques for determining sample size. Two that are of primary importance are the following:

- Whether the technique deals with fixed or sequential sampling
- Whether its logic is based on traditional or Bayesian inferential methods

Other than the brief discussion of sequential sampling that follows, this chapter is concerned with the determination of a fixed sample size with emphasis on traditional inference, such as Neyman-Pearson, rather than Bayesian inference.[1]

Although the discussion will focus on the statistical aspects of setting sample size, it should be recognized that nonstatistical dimensions can affect the value of a research project. Such things as study objectives, the length of a questionnaire, budget, time schedule, and the requirements of data analysis procedures, all have a direct effect on sample size decisions.

Fixed Versus Sequential Sampling

As the name implies, in fixed-size sampling the number of items sampled is decided in advance. The size of the sample is chosen to achieve a balance between sample reliability and sample cost. In general, all sample observations are taken before the data are analyzed.

In sequential sampling, however, the number of items is not preselected. Rather, the

Table 7.3 Proportion Based Confidence Intervals Computed at the 95% Confidence Level

Sample Size	Variability Proportions					
	50/50%	40/60%	30/70%	20/80%	90/10%	95/5%
25	20	19.6	18.3	16	12	8.7
50	14.2	13.9	13	11.4	8.5	6.2
75	11.5	11.3	10.5	9.2	6.9	5
100	10	9.8	9.2	8	6	4.4
150	8.2	8	7.5	6.6	4.9	3.6
200	7.1	7	6.5	5.7	4.3	3.1
250	6.3	6.2	5.8	5	3.8	2.7
300	5.8	5.7	5.3	4.6	3.5	2.5
400	5	4.9	4.6	4	3	2.2
500	4.5	4.4	*4.1	3.6	2.7	2
600	4.1	4	3.8	3.3	2.5	1.8
800	3.5	3.4	3.2	2.8	2.1	1.5
1000	3.1	3.0	2.8	2.5	1.9	1.4
1500	2.5	2.5	2.3	2.0	1.5	1.1
2000	2.2	2.2	2.0	1.6	1.2	0.96
2500	2	1.9	1.8	1.6	1.2	0.85
5000	1.4	1.4	1.3	1.1	.83	0.6

*Example Interpretation: In a product usage study where the expected product usage incidence rate is 30%, a sample of 500 will yield a precision of +/- 4.1 percentage points at the 95% confidence level.

This table is computed using the following formula:

$$\frac{(\text{Number of Standard Errors})^2 * ((\text{proportion})*(1\text{-proportion})) / (\text{Accuracy})}{(1+((\text{Number of Standard Errors})^2 * ((\text{proportion})*(1\text{-proportion})) / (\text{Accuracy})\text{-}1) / (\text{the population size})}$$

This formula is easily entered into a spreadsheet, to compute a sample size determination table.

Sampling Basics: Sampling Distributions and Standard Errors

The reader will recall from elementary statistics the concept of a sampling distribution. For a specified sample statistic (e.g., the sample mean) the sampling distribution is the probability distribution for all possible random samples of a given size *n* drawn from the specified population. The standard error of the statistic is the standard deviation of the specified sampling distribution. We shall use the following symbols in our brief review of the elementary formulas for calculating the standard error of the mean and proportion (under simple random sampling):

μ = population mean
π = population proportion regarding some attribute
σ = standard deviation of the population
s = standard deviation of the sample, adjusted to serve as an estimate of the standard deviation of the population
$\overline{X}$ = arithmetic mean of a sample
p = sample proportion
n = number of items in the sample

7. If the population standard deviation σ is not known, which is often the case, we can estimate it from the sample observations by use of the following formula:

$$s = \sqrt{\frac{\sum_{i=1}^{n}(X_1 - \overline{X})^2}{n-1}}$$

We can consider s to be an *estimator* of the population standard deviation, σ. In small samples (e.g., less than 30), the t distribution of Table A.2 in Appendix A is appropriate for finding probability points. However, if the sample size exceeds 30 or so, the standardized normal distribution of Table A.1 is a good approximation of the t distribution.
In cases where σ is estimated by s, the standard error of the mean becomes

$$est.\sigma_{\overline{X}} = \frac{s}{\sqrt{n}}$$

where $est.\sigma_{\overline{X}}$ denotes the fact that σ is estimated from s, as defined in the preceding equation.

8. Analogously, in the case of the standard error of the proportion, we can use the sample proportion p as an estimate of π to obtain

$$est.\sigma_{\overline{P}} = \sqrt{\frac{p(1-p)}{n}}$$

as an estimated standard error of the proportion. Strictly speaking, $n-1$ should appear in the denominator. However, if n exceeds about 100 (which is typical of the samples obtained in marketing research), this adjustment makes little difference in the results.

Methods of Estimating Sample Size

In our discussion of sample planning we pointed out that there are four traditional approaches to determining sample size.

- First, the analyst can simply select a size either arbitrarily or on the basis of some judgmentally based criterion. Similarly, there may be instances where the size of sample represents all that was available at the time—such as when a sample is composed of members of some organization and data collection occurs during a meeting of the organization.

- Second, analysis considerations may be involved and the sample size is determined from the minimum cell size needed. For example, if the critical aspect of the analysis required a cross-tabulation on three categorical variables that created 12 cells (2 categories x 3 categories x 2 categories = 12 cells), and it was felt that there should be at least 30 observations in a cell, then the absolute minimum sample size needed would be 360.

- Third, the budget may determine the sample size. If, for example, the research design for an online survey of physicians, the cost of each interview was estimated to be $50, and the budget allotted to data collection was $10,000, then the sample size would be 200.

It may appear that these methods are for nonprobability samples. While this certainly is true, these methods are also applicable to probability samples and have occasionally been used

$$n = \frac{N\pi(1-\pi)Z^2}{NE^2 + \pi(1-\pi)Z^2}$$

Determining Sample Size When More Than One Interval Estimate Is to Be Made from the Same Sample

The usual case when collecting sample data for estimation of various parameters is that more than one estimate is to be made. The sample size for each of the estimates will usually be different. Since only one sample is to be chosen, what size should it be?

A strict adherence to the allowable error and the confidence levels specified in the calculation of the sample sizes for the individual estimation problems leaves no choice but to take the largest sample size calculated. This will give more precision for the other estimates than was specified but will meet the specification for the estimate for which the size of sample was calculated. In the specialty fruit/energy drink consumption problem, for example, the sample size would be 369 (the sample size calculated for estimating the proportion of users) rather than 246 (the sample size calculated for estimating the mean amount used). Remember, the sample sizes determined in this manner are for obtained samples. In order to determine the size of the original sample the researcher must estimate the rate of response expected. For example, if a mail survey was to be conducted and it was believed that only a 20% response would be obtained, a desired obtained sample of 250 would need an original sample of 1,250.

Devices for Calculating Sample Size

In practice, manual devices (paper nomographs) and online sample size calculators can be used to easily compute a sample size helpful in rough-guide situations where the researcher is not sure of either allowable error levels or population standard deviations. One such calculator is located at http://marketing.byu.edu/samplesizecalculator.html .

The Hypothesis-Testing Approach

As indicated earlier, sample sizes can also be determined (within the apparatus of traditional statistical inference) by the hypothesis-testing approach. In this case the procedures are more are more elaborate. We shall need both an assumed probability of making a type I error—called the alpha risk—and an assumed probability of making a type II error—called the beta risk. These risks are, in turn, based on H_0: the null hypothesis, and H_1: the alternate hypothesis.

In hypothesis testing the sample results sometimes lead us to reject H_0 when it is true. This is a type I error. On other occasions the sample findings may lead us to accept H_0 when it is false. This is a type II error. The nature of these errors is shown in Table 7.4.

A numerical example should make this approach clearer. We first consider the case for means and then the case for proportions. Before doing this, however, a few words on the relationship between the Type I and Type II errors are in order. The relationship between these two errors is an inverse. The ability of a sample to protect against the type II error is called statistical power.

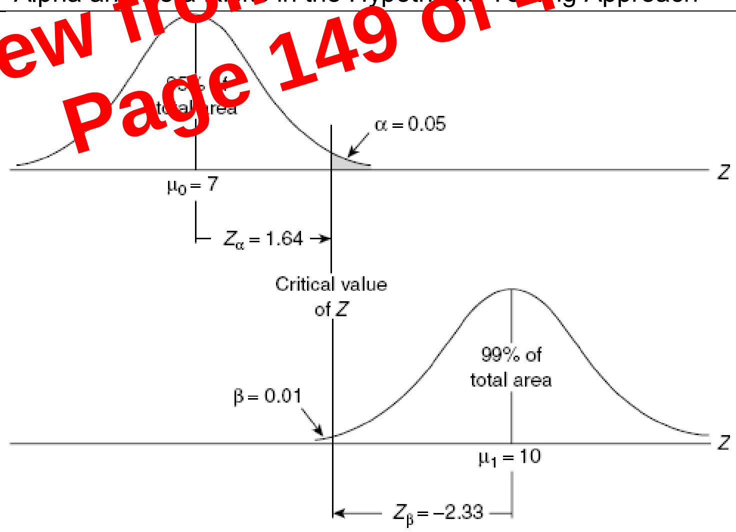
The Case Involving Means

As an illustrative example, let us assume that a store test of a new bleaching agent is to be conducted. It has been determined earlier that if the (population) sales per store average only 7 cases per week, the new product should not be marketed. On the other hand, a mean sales level of 10 cases per week would justify marketing the new product nationally. Using methods of traditional inference, how should the number of sample stores for the market test be determined? The procedures are similar to those for interval estimation problems but are somewhat more complicated. Specifically, we go through the following checklist:

1. Specify the values for the null (H_0) and the alternate (H_1) hypotheses to be tested in terms of population means μ_0 and μ_1, respectively. (By convention, the null hypothesis is the one that would result in no change being made, if accepted.) In the bleach market-introduction problem, the values are set at H_0: μ_0 = 7 cases per week, and H_1: μ_1 = 10 cases per week.

2. Specify the allowable probabilities (α and β, respectively) of type I and type II errors. The type I error is the error of rejecting a true null hypothesis. The type II error is made when the alternate hypothesis is rejected when it is true. α and β are the allowable probabilities of making those two types of errors, respectively. They are shown graphically in Figure 7.3, where we assume that in the bleach-introduction problem the allowable probabilities of error are assigned as α = 0.05 and β = 0.01.

Figure 7.3 Alpha and Beta Risks in the Hypothesis-Testing Approach



3. Determine the number of standard errors associated with each of the error probabilities α and β. For a one-tailed test the Z values for the 0.05 and 0.01 risks, respectively, are found from Table A.1 in Appendix A to be Z_α = 1.64 and Z_β = 2.33. These are shown in figure 7.3. Note that in the figure we affix a minus sign to the value of Z_β since the critical values lies to the left of μ_1 = 10.

CHAPTER 8
EXPERIMENTATION

Experimentation is widely used in marketing research. Marketing experiments have been conducted in such diverse activities as evaluating new products, selecting advertising copy themes, determining the frequency of salespeople's calls, and evaluating all aspects of a movie (including ending, pacing, music, and even the story line). For example, as shown in Exhibit 8.1 the ending of the very successful movie Fatal Attraction was changed because test audiences did not like the original ending.

This chapter discusses the objectives of experimentation illustrates techniques for designing and analyzing marketing experiments including:

- the nature of experimentation
- ingredients of a marketing experiment
- sources of invalidity
- models of experimental design
- panels and experimental design
- difficulties in field experiments in marketing

This chapter further discusses experimental designs and logic for advanced surveys using the capabilities of Qualtrics.com, including the following:

- piping of text, graphics and experimental treatments
- simple branching and compound branching logic, looping and piping of answers
- question and treatment blocks
- quota fulfillment
- randomization of answer choices, questions, treatment blocks and alternative questionnaire forms.

The Nature of Experimentation

Two general types of experimental designs exist—natural and controlled. A natural experiment is one in which the investigator intervenes only to the extent required for measurement, and there is no deliberate manipulation of an assumed causal variable. “Nature” produces the changes. In contrast, in a controlled experiment two kinds of intervention are needed:

1. manipulation of at least one assumed causal variable
2. random assignment of subjects to experimental and control groups.

True experiments have both types of intervention, while *quasi-experiments* manipulate the variables but do not randomly assign the subjects. All true experiments have certain things in common—treatments (i.e., assumed causal variables), an outcome measure, units of assignment, and some comparison from which change can be inferred and, it is hoped, attributed to the treatment. Quasi-experiments, on the other hand, have treatment, outcome measures, and experimental units but do not randomly assign subjects to treatments. Rather, subjects already

Factorial Designs

A factorial experiment is one in which an equal number of observations is made of all variable combinations. The variables must have at least two levels or categories each. In essence, the factorial design is one that has combined two or more completely randomized designs into a single experiment. This type of experiment enables the researcher to study possible interactions among the variables of interest. Suppose we return to our Ad illustration but now assume that the researcher is interested in studying the effects of two variables of interest: Ad Spokesperson type (at three levels – Male, Female, Cartoon Character) and Message Type (at two levels – Benefits Message and Humorous Message). The design is shown in Table 8-2. Note that each combination of M_iS_j occurs only once in the design. While the plan still is to use a single sample for the experiment, the researcher intends to randomize the presentation of each combination among the various respondents (each receiving one Ad).

In the factorial experiment we can test for all main effects (i.e., spokesperson types, messages), and in this case, where we have replicated each combination, for the interaction of the variables as well. If the interaction term is significant, ordinarily the calculation of main effects is superfluous, since the experimenter will customarily be interested in the best combination of variables. That is, in market experimentation the researcher is typically interested in the combination of controlled variables that leads to the best payoff in terms of preference, sales, market share, cash flow, or some other measure of effectiveness.

Table 8-2 Factorial Design—Advertising Experiment

	Ad Spokesperson Gender		
Message	*Male*	*Female*	*Cartoon Character*
Benefit Statements	M_1S_1	M_1S_2	M_1S_3
Humorous Statements	M_2S_1	M_2S_2	M_2S_3

Latin Square

Latin-square designs are multivariable designs that are used to reduce the number of observations that would be required in a full factorial design. In using Latin-square designs the researcher is usually assuming that interaction effects are negligible; in so doing, all main effects can be estimated by this procedure.

As an illustration of a Latin-square design, suppose that the researcher is interested in three variables (each at four levels) on store sales. For example, the researcher may be interested in the effect of shelf placement on the sale of energy drinks and be able to manipulate the following variables for product placement:

A: shelf height—four levels—knee level, waist level, eye level, reach level
B: shelf facings—four levels—25%, 50%, 75%, and 100% of total width of gondola
C: shelf fullness—four levels—25%, 50%, 75%, and 100% of total height of gondola section

If the researcher were to run a full factorial experiment, with one replication only, there would be (4)3 = 64 observations required. By using a Latin-square design only 16 observations are required (with estimation of main effects only).

subject. Suppose that it is too costly to screen subjects, and only those with approximately the same intelligence quotient are selected. We shall assume that the researcher is able to measure each subject's IQ.

In this type of situation, the researcher may use covariance analysis. Roughly speaking, the computational procedure is similar to a regression problem. The researcher, in effect, determines the effect on response resulting from differences (in IQ) among test units and removes this influence so that the effect of the controlled variables can be determined independently of the effect of test differences on response.

Recapitulation

As noted earlier, the study of experimental design is basically the study of two things:

1. various experimental layouts, such as single factor, factorial, Latin-square, and randomized block designs
2. analysis of variance and covariance techniques for testing whether the various treatment effects are significant

Many other kinds of design layouts, including fractional factorial designs, balanced incomplete blocks, hierarchical designs, split-plot designs, and partially balanced incomplete blocks are available. We have only described briefly the characteristics of some of the basic statistical designs.

We have not covered the issue of how to select an experimental design. The design chosen must be congruent with the research question and capable of implementation with resources that are available. Such issues as the number of independent variables, the sources and number of extraneous variables, the nature of the dependent variable, the number of subjects available for participation in the experiment and other methodological issues will have a bearing on what might be an appropriate design to use. Increasingly, marketing researchers use personal computer software to aid in this selection.

The Panel as a Natural Experimental Design

The normal course of operation of a consumer panel generates a continuing set of natural experimental data. Customer responses to changes in any of the controllable or environmental variables affecting purchase decisions are recorded in the normal process of conducting the panel. Audience and deal panels provide similar response measurements.

Time-series, cross-sectional, and combination cross-sectional, time-series designs are all inherent in panel data. To illustrate their application, suppose that we have increased the price of our product in selected territories. We can analyze the price-increase effect, at either the aggregated or individual household level, using the data from those territories in which price was increased with either the after-only without control group or the before-after without control group designs [classical designs (1) and (2)]. A cross-sectional analysis may be made by comparing, for a given period after the increase, the purchase data for the territories in which the price was raised with those in which no change was made [classical design (4)]. A preferable approach here would be to use a combination cross-sectional, time-series design and compare the change in purchases before and after the price increase in the territories in which price was not changed (control group). Such a study could employ either classical design (3) or (5).

The limitations of each of these designs discussed earlier apply when they are used with panel data as well. A major difficulty, or course, is in sorting out the effect of the price increase from the extraneous producers affecting purchases over time and among territories. In this illustration selective price increases by territory would only have been made in response to been added into the Block Randomizer. The number in the dropdown reflects the number of blocks you want the respondents to see. You may set it to show all of the blocks in the randomizer, or a subset of those blocks (i.e., If 4 blocks are added to the randomizer and each respondent is to randomly respond to two of those blocks, put a "2" in the box.

Randomization using Multiple Questionnaires

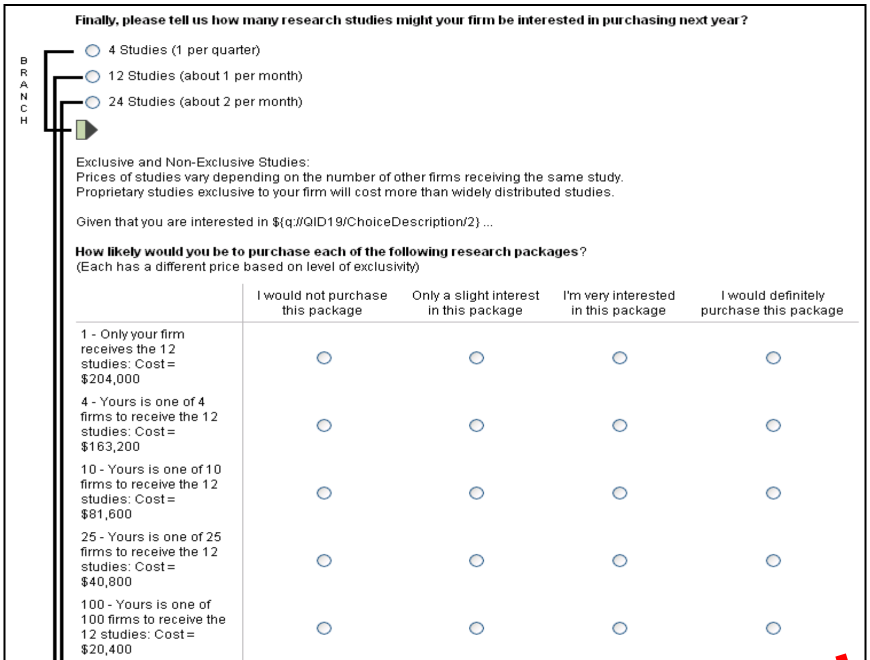
Occasionally, experimental treatments are so extensive that a completely different questionnaire is required for each treatment. In such cases, the questionnaires are prepared and randomly assigned to the respondents. Such a design may be achieved by either using a "portal survey" where the links to the other surveys are randomly presented via questions or block randomization, or by inserting a separate java script into a question. The java script would be similar to the one that follows to randomly assign respondents to a survey.

Sample Experimental Designs and Associated Survey Flows

Script to Open Survey in the Same Window, or in a Separate Window:

```
<SCRIPT LANGUAGE="JAVASCRIPT" TYPE="TEXT/JAVASCRIPT">
Event.observe(window, 'load', function() {
        <!-- Javascript to redirect user to a randomly selected survey //-->
        // Array structure - as many links as you need
        var seedcount = 7
        var index = 3;
        var linkArray = new Array(index);
        // Load array with available survey links here:
        linkArray[0] = "http:// new.qualtrics.com/SE?SID=SV_6LR6QjjYiHkSe21&SVID=Prod";
        linkArray[1] = "http:// new.qualtrics.com/SE?SID=SV_6LR6QjjYiHkSe22&SVID=Prod";
        linkArray[2] = "http:// new.qualtrics.com/SE?SID=SV_6LR6QjjYiHkSe23&SVID=Prod";
        // Generate random number that will correspond to an index in the above array.
        // The integer in the equation is the total number of links you have minus 1
        now = new Date();
        var seed = now.getSeconds();
        randNum = Math.round(1000000*Math.random(seed))
        for (loopcount = 0; loopcount < seedcount; loopcount++) {
          randNum += Math.round(1000000*Math.random(randNum * seed));
        }
        randNum = Math.round(1000000*Math.random(randNum)) % index;
        // Use one of the two statements below by changing the comment
        // Redirect user to appropriate random survey page in the current window
         window.location.replace(linkArray[randNum]);
        // Or open the appropriate random survey page in a new window
        // window.open(linkArray[randNum]);
});
</script>
```

Figure 3 Number of Studies x Exclusivity for Branch (Cell 1-1 of the Design – Price Matrix)



Finally, please tell us how many research studies might your firm be interested in purchasing next year?

B
R
A
N
C
H

- 4 Studies (1 per quarter)
- 12 Studies (about 1 per month)
- 24 Studies (about 2 per month)

Exclusive and Non-Exclusive Studies:
Prices of studies vary depending on the number of other firms receiving the same study.
Proprietary studies exclusive to your firm will cost more than widely distributed studies.

Given that you are interested in ${q://QID19/ChoiceDescription/2} ...

How likely would you be to purchase each of the following research packages?
(Each has a different price based on level of exclusivity)

	I would not purchase this package	Only a slight interest in this package	I'm very interested in this package	I would definitely purchase this package
1 - Only your firm receives the 12 studies: Cost = $204,000	○	○	○	○
4 - Yours is one of 4 firms to receive the 12 studies: Cost = $163,200	○	○	○	○
10 - Yours is one of 10 firms to receive the 12 studies: Cost = $81,600	○	○	○	○
25 - Yours is one of 25 firms to receive the 12 studies: Cost = $40,800	○	○	○	○
100 - Yours is one of 100 firms to receive the 12 studies: Cost = $20,400	○	○	○	○

decisions, since an association was found to exist between families with these characteristics and the purchase of color televisions sets.

In studies of consumers where there is a basis for believing that such associations might exist, researchers obtain information on one or more socioeconomic characteristics; those most frequently obtained are income, occupation, level of education, age, sex, marital status, and size of family. While socioeconomic characteristics are by far the most widely used bases for classification of consumers, other bases exist. Among these are attitudes, preferences, personality traits, perceived risk, and such measures of actual buying behavior as amount purchased and brand loyalty. It may be interesting to know, for example, that owners of SUVs show different personality traits than owners of other vehicles; such knowledge will be useful in marketing automobiles, however, only if it can be used to develop and evaluate appeals for each type of buyer. Doing so can enhance segmentation, positioning, and market targeting.

In general, the identification of consumer segments is useful in marketing so long as the following four statements apply:

1. **Substantial**: The value in terms of potentially increased sales makes it worthwhile to do so.
2. **Differentiable**: There are practical means of differentiating purchase behavior among market segments. There is homogeneity within and heterogeneity between segments.
3. **Operational**: There is a cost effective means of reaching the targeted market segment
4. **Responsive**: The differentiated market segments respond differentially to marketing offerings tailored to meet their needs.

Two commonly used and widely accepted classifications of consumers are by stage of the life cycle and by lifestyle. One classification identifies the household life-cycle groups as the following:

1. Young unmarrieds
2. Young marrieds, no children
3. Young marrieds, with children, youngest child under six
4. Older marrieds, with children, youngest child six or older
5. Older marrieds, with children maintaining separate households
6. Solitary survivors or older single people

Some writers have expanded the number of stages by distinguishing in the last two stages whether a person is in the labor force or retired. See Wells and Gubar (1966) and Wagner and Hanna (1983) for more detailed explanations of the life-cycle concept and marketing research.

The life-cycle stage has obvious implications with respect to purchases associated with family formation (furniture, appliances, household effects, and housing) and addition of children (food, clothing, toys, expanded housing). Other, less obvious relationships exist as well. New-car buying reaches its peak among the older married couples whose children have passed the age of six. A second stage of furniture buying takes place when children begin to date and have parties at home. Dental work, travel, and purchases of insurance are examples of service purchases associated with the life cycle.

Lifestyle has a close association with membership in a social class. It is a basis for segmenting customers by values, activities, interests and opinions, as well as by income. These differences tend to be expressed through the products bought and stores patronized, as well as the area in which one lives, club membership, religious affiliation, and other means. The media used for expression are often either consciously or subconsciously, symbolic representations of the class to which the person perceives he or she belongs (or would like to belong). When used with personality traits, lifestyle variables form the basis of psychographic classification, as illustrated in Exhibit 9.2. An illustration of psychographic questions is shown in Table 9.3.

As an example, let us consider the life styles of the Harley Owners Group (HOG). By examining a group of questions used in a segmentation study of values and motorcycle use, we find a divergent group of lifestyles that have embraced the mystique of owning a Harley.

EXHIBIT 9.2 Harley Owners Group (HOG) Classification by Psychographics

Psychographic research has suggested many different segmentation schemes. Such schemas represent interesting demographic and product markets, and provide a much more colorful description of the group as a whole as well as the diversity within.

Research by William Swinyard (1994a, 1994b) suggests that Harley-Davidson owners are a diverse group consisting of six distinct segments with very different motorcycling lifestyles:

- **Tour Glides** find the appeal of motorcycling in long distance touring. They like riding long distances, use their bike both for touring and everyday transportation, are more interested in the comfort of their motorcycle than its speed, prefer riding with a passenger, and wear a helmet.

  More than the average Harley rider, Tour Glides are religiously traditional, have somewhat old-fashioned tastes and habits, are disciplinarians with their children, like reading, and feel they live a full and interesting life. They are less ambitious than others, and are distinctively unattracted by social gatherings and danger.

- **Vanilla Dream Riders.** The Vanilla Dream Riders are more interested in the dream of motorcycling than in motorcycling itself, and are otherwise just plain vanilla—a relatively undistinguished group.

  This is the largest, oldest, wealthiest, and among the best educated segment of Harley owners, who have the newest motorcycles yet ride them least, and spend little in accessorizing them. You see the Dream Riders taking riding on short trips around town (often by themselves), wearing a helmet, and riding a stock bike. They are distinctively unaffiliated with the "live to ride" ethic, and receive relatively little psychic satisfaction from riding. Their motorcycle is merely a possession, having no real place as a "family member." They are conservative in their moral values, marital roles, and daily behavior.

- **The Hard Core** segment is on the fringe of society, and identifies with the stereotypical biker subculture.

  They are the youngest, next-to-least well-educated, and certainly the poorest, yet spend nearly 50 percent more than any other segment on accessorizing their motorcycles. Virtually all are blue-collar workers. In relative terms, Hard Core members are much more likely than others to feel like an outlaw, and believe

If you were offered both types of mortgages, indicate the difference, if any, between the interest rate for the fixed-rate plan and the initial interest rate for the variable-rate plan.

Fixed rate was higher	No difference	Variable rate was higher	Cannot recall	Did not inquire
○	○	○	○	○

Attitudes and Opinions

Investigators in the behavioral science fields of psychology, sociology, and political science have made extensive studies of attitudes and opinions over a wide range of subject areas. The study of people's behavior in business and economic contexts is also a behavioral science. As such, it has been a natural consequence that marketing research has adopted, adapted and applied many techniques, including attitude-opinion studies to obtain information applicable to the solution of marketing problems.

The terms *attitude* and *opinion* have frequently been differentiated in psychological and sociological investigations. A commonly drawn distinction has been to view an attitude as a predisposition to act in a certain way, and an opinion as a verbalization of an attitude. Thus, a statement by a respondent that he or she prefers viewing Blue-Ray DVD's to HDTV color television programs would be an opinion expressing (one aspect of) the respondent's attitude toward high-definition television.

When used to predict actions that the respondent will take, this distinction between attitude and opinion rapidly becomes blurred. Since the major purpose of attitude-opinion research in marketing is to predict behavior, this differentiation is, at best, of limited usefulness. We shall therefore use the terms interchangeably. Attitude research in marketing has been conducted with the use of both qualitative and quantitative techniques. In either form, researchers encounter problems that are more severe than those involved in obtaining any other type of descriptive information discussed. Despite these problems, which we will discuss in later chapters in some detail, attitude-opinion research has been widely used to provide information for choosing among alternatives. Its greatest use has been in the areas of product design (including packaging and branding) and advertising. Other uses have been in selecting store locations, developing service policies, and choosing company and trade names. In fact, attitudes and opinions are central in customer satisfaction studies.

Measuring Attitudes

Expectancy Value Measures of Behavioral Intention (BI) and Attitudes (A)

Expectancy value models were first developed in the 1960's as a method of predicting behavioral intentions (a precursor of actual behavior). Expectancy value models use attitudes to predict behavioral intention (intention to try, purchase, recommend, or re-purchase a product or service). This methodology has become a mainstay of marketing research and is found to perform well in predicting both consumer behavior and consumer satisfaction/dissatisfaction.

The Expectancy value model uses attitudes and beliefs in a mathematical formulation that is read as follows:

Exhibit 9.4 Common Ingredients of a Customer Satisfaction Survey

Product Use
Frequency of product use
Primary use location
Primary precipitating events or situations for product use or need
Usage rates and trends

Product Familiarity
Degree of actual product use familiarity
Knowledge (read product information, read product label, etc.)
Knowledge and Involvement with product and the purchase process
Awareness of other brands
Reasons for original product purchase (selection reasons)
Primary benefits sought from the product

Product Evaluation
Attribute evaluation matrix: (quality, price, trust, importance, performance, value)
Perceived benefit associations matrix
Importance, performance
Identification of primary benefits sought
Comparison to other brands (better, worse)
What is the best thing about the brand, what could be done better

Message and Package Evaluation
Packaging size, design
Advertising Promise, message fulfillment evaluation

Value Analysis
Expectation of price
Expectation of relative price (full price, on sale)
Current price paid

Satisfaction Measurements
Overall Satisfaction
Reasons for Satisfaction Evaluation
Satisfaction with attributes, features, benefits
Satisfaction with use
Expected and Ideal Satisfaction-Performance Measures

Likelihood of recommending
Likelihood of repurchasing

What Is Customer Satisfaction?

Customer satisfaction measures indicate how well a company's products or services meet or exceed customer expectations. These expectations will reflect many aspects of the company's business activities including the actual product, service and company. Customer satisfaction measures will tap the customer's lifetime of product and service experience.

Interval Scales

Interval scales possess a constant unit of measurement and permit one to make meaningful statements about differences separating two objects. This type of scale possesses the properties of order and distance, but the zero point of the scale is arbitrary. Among the most common examples of interval scaling are the Fahrenheit and Centigrade scales used to measure temperature, and various types of indexes like the Consumer Price Index. While an arbitrary zero is assigned to each temperature scale, equal temperature differences are found by scaling equal volumes of expansion in the liquid used in the thermometer. Interval scales permit inferences to be made about the differences between the entities to be measured (say, warmth), but we cannot meaningfully state that any value on a specific interval scale is a multiple of another.

An example should make this point clearer. It is not empirically correct to say that 50°F is twice as hot as 25°F. Converting from Fahrenheit to Centigrade, we find that the corresponding temperatures on the centigrade scale are 10°C and –3.9°C, which are not in the ratio 2:1. We can say, however, that differences between values on different temperature scales are multiples of each other. That is, the difference of 50°F–0°F is twice the difference of 25°–0°F. The corresponding differences on the Centigrade scale are 10°C – (–17.7°C) = 27.7°C and –3.9°C – (–17.7°C) = 13.8°C are in the same 2:1 ratio.

Interval scales are unique up to a transformation of the form $y = a + bx$; $b > 0$. This means that interval scales can be transformed from one to another by adding or multiplying a constant. For example, we can convert from a Fahrenheit to Celsius using the formula:

$$T_C = 5/9\ (T_F - 32)$$

Most ordinary statistical measures (such as arithmetic mean, standard deviation, and correlation coefficient) require only interval scales for their computation.

Ratio Scales

Ratio scales represent the elite of scales and contain all the information of lower-order scales and more besides. These are scales like length and weight that possess a unique zero point, in addition to equal intervals. All types of statistical operations can be performed on ratio scales.

An example of ratio-scale properties is that 3 yards is three times 1 yard. If transformed to feet, then 9 feet and 3 feet are in the same 3:1 ratio. It is easy to move from one scale to another merely by applying an appropriate positive multiplicative constant; this is the practice followed when changing from grams to pounds or from feet to inches.

Relationships Among Scales

To provide some idea of the relationships among nominal, ordinal, interval, and ratio scales, the marketing researcher who uses descriptive statistics (arithmetic mean, standard deviation) and tests of significance (t-test, F-test) should require that the data are (at least) interval-scaled .

From a purely mathematical point of view, you can obviously do arithmetic with any set of numbers—and any scale. What is at issue here is the interpretation and meaningfulness of the results. As we select more powerful measurement scales, our abilities to predict, explain, and otherwise understand respondent ratings also increase.

Table 10.1 Scales of Measurement

Scale	*Mathematical Group Structure*	*Permissible Statistics*	*Typical Elements*
Nominal	Permutation group y = *f(x)*, where *f(x)* means any one-to-one correspondence	Mode Contingency Coefficient	Numbering of football players Assignment of type or model numbers to classes
Ordinal	Isotonic group y = *f(x)*, where *f(x)* means any strictly increasing function	Median Percentile Order correlation Sign test; run test	Hardness of minerals Quality of leather, lumber, wool, etc. Pleasantness of odors
Interval	General linear group $y = a+bx$ $b > 0$	Mean Average deviation Standard deviation Product-moment correlation *t*-test, F-test	Temperature (Fahrenheit and centigrade) Energy Calendar dates
Ratio	Similarity group $y = cx$ $c > 0$	Geometric mean Harmonic mean Coefficient of variation	Length, width, density, resistance Pitch scale, loudness scale

(Stevens 1946, p. 678)

Basic Question and Answer Formats

Underlying every question is a basic reason for asking it. This reason reflects the construct to be measured, the problem to be solved or hypothesis to be tested. Constructing a question that reflects this reason will result in a higher probability that the desired response will be obtained. Table 10.2 shows nine different types of questions (based on the nature of content), the broad reason underlying asking each type of question, and some examples of each type.

Table 10.2 Basic Question Types

Type of Question	*Goal of Question*	*Positioning of Question*
Factual or behavioral	To get information.	Questions beginning with what, where, when, why, who and how.
Explanatory	To get additional information or to broaden discussion.	How would that help? How would you go about doing that? What other things should be considered?
Attitudinal	To get perceptions, motivations, feelings, etc., about an object or topic.	What do you believe to be the best? How strongly do you feel about XYZ?
Justifying	To get proof to challenge old ideas and to get new ones.	How do you know? What makes you say that?
Leading	To introduce a thought of your own.	Would this be a possible solution? What do you think of this plan?
Hypothetical	To use assumptions or suppositions.	What would happen if we did it this way? If it came in blue would you buy it today?
Alternative	To get a decision or agreement.	Which of these plans do you think is best? Is one or two o'clock best for you?
Coordinative	To develop common agreement. To take action.	Do we all agree that this is our next step?
Comparative	To compare alternatives or to get a judgment anchored by another item.	Is baseball more or less exciting to watch on TV than soccer?

with respect to a long list of attributes. In an online survey environment, respondents can quickly scan down columns or across screens and quickly complete the pick data task for a familiar brand, thereby saving time and reducing respondent fatigue and dropout rates.

Having people describe a brand by picking attributes from a list is a quick and simple way to assess brand performance and positioning. Whitlark and Smith (2004) show that when respondents are asked to pick from one third to one half of the viewed items, the pick *k of n* data can be superior to scaled data in terms of reliability and power to discriminate between attributes.

Rank-Order Questions/Answers

The next level of measurement rank-orders the answers and thereby increases the power of the measurement scale over categorical measurement by including the characteristic of order in the data. Whereas the categorical data associated with many dichotomous or multiple-choice items does not permit us to say that one item is greater than another, rank-order data allows for the analysis of differences. Rank-order questions use an answer format that requires the respondent to assign a rank position to all items, or a subset of items in the answer list. The first, second, and so forth up to the *n*th item would be ordered. Procedures for assigning position numbers can be very versatile, resulting in different types of questions that can be asked. Typical questions might include identifying preference rankings, or attribute associations from first to last, most recent to least recent or relative position (most, next most, and so forth, until either a set number of items is ordered or all items may be ordered).

When this type of question is administered online or using a CATI (Computer Aided Telephone Interviewing) system, additional options for administration may include randomization and acceptance/validation of ties in the ranking. Randomization of the answer list order helps to control for presentation order bias. It is well established that in elections, being the first in a ballot candidate list increases chances of receiving the voter's selection.

Tied rankings are another issue to be considered for rank-order questions. When ties are permitted, several items may be evaluated as having the same rank. In general, this is not a good idea because it weakens the data. However, if ties truly exist, then the ranking should reflect this. Rank-order questions are generally a difficult type of question for respondents to answer, especially if the number of items to be ranked goes beyond five or seven.

Constant Sum Questions/Answers

A constant sum question is a powerful question type that permits collection of ratio data, meaning that the data is able to express the relative value or importance of the options (option A is twice as important as option B). This type of question is used when you are relatively sure of the answer set (i.e., reasons for purchase, or you want to evaluate a limited number of reasons that you believe are important). The following example of a constant sum question from Qualtrics, uses sliding scales to select a sum of 100 points:

ADVANCED MEASUREMENT AND SCALING CONCEPTS

Continuing our discussion on scales, we now focus on some of the more common scaling techniques and models. We focus on broad concepts of attitude scaling—the study of scaling for the measurement of managerial and consumer or buyer perception, preference, and motivation. All attitude (and other psychological) measurement procedures are concerned with having people—consumers, purchasing agents, marketing managers, or whomever—respond about certain stimuli according to specified sets of instructions. The stimuli may be alternative products or services, advertising copy themes, package designs, brand names, sales presentations, and so on. The response may involve judging which copy theme is more pleasing than another, which package design is more appealing than another, what mental images do new brand names evoke, which adjectives best describe each salesperson, and so on.

Scaling procedures can be classified in terms of the measurement properties of the final scale (nominal, ordinal, interval, or ratio), the task that the subject is asked to perform, or in still other ways, such as whether the scale measures the subject, the stimuli, or both (Torgerson, 1958).

We begin with a discussion of various methods for collecting ordinal-scaled data (paired comparisons, rankings, ratings, etc.) in terms of their mechanics and assumptions regarding their scale properties. Then specific procedures for developing these actual scales are discussed. Techniques such as Thurstone Case V scaling, semantic differential, the Likert summated scale, and the Thurstone differential scale are illustrated. The chapter concludes with some issues and limitations of scaling.

Advanced Ordinal Measurement Methods

The variety of ordinal measurement methods includes a number of techniques:

- Paired comparisons
- Ranking procedures
- Ordered-category sorting
- Rating techniques

We discuss each of these data collection procedures in turn.

Paired Comparisons

As the name suggests, paired comparisons require the respondent to choose one of a pair of stimuli that "has more of", "dominates", "precedes", "wins over", or "exceeds" the other with respect to some designated property of interest. If, for example, six laundry detergent brands are to be compared for "sudsiness", a full set of paired comparisons would involve (n x n – 1)/2 = (6 × 5) / 2, or 15, paired comparisons (if order of presentation is not considered). Respondents are asked which one of each pair has the most sudsiness.

A sample question format for paired comparisons is shown in Table 10.4. The order of presentation of the pairs and which item of a pair is shown first are typically determined and/or presented randomly. Consider the following hypothetical brand names (and numerical categories): Arrow (1), Zip (2), Dept (3), Advance (4), Crown (5), and Mountain (6).

Table 10.4 Example of the Paired Comparisons Question

For each of the pairs of laundry detergent brands shown below, indicate which one has the most sudsiness:

I prefer the left Brand		I prefer the right Brand
	Arrow ----- Zip	
○		○
	Arrow ----- Advance	
○		○
	Dept ----- Arrow	
○		○
	Crown ----- Arrow	
○		○
	Crown ----- Mountain	
○		○

Original Data

		Brand					
		Arrow	Zip	Advance	Dept	Crown	Mountain
Brand	Arrow	X	0	1	1	1	1
	Zip	1	X	1	1	1	1
	Advance	0	0	X	0	0	0
	Dept	0	0	1	X	0	0
	Crown	0	0	1	1	X	1
	Mountain	0	0	1	1	0	X

or ratio-scaled. The flexibility of rating procedures makes them appropriate for either the ordinal or interval/ratio measurement data collection methods (depending on the nature of the scale values).

The rating task typically involves having a respondent place that which is being rated (a person, object, or concept) along a continuum or in one of an ordered set of categories. Ratings allow the respondent to designate a degree or an amount of a characteristic or attribute as a point on a scale. The task of rating is one of the most popular and easily applied data collection methods, and is used in a variety of scaling approaches, such as the semantic differential and the Likert summated scale.

Rating scales can be either monadic or comparative. In monadic scaling, each object is measured (rated) by itself, independently of any other objects being rated. In contrast, comparative scaling objects are evaluated in comparison with other objects. For example, a recent in-flight survey conducted by United Airlines asked the following questions:

Monadic Example:
Please rate the service you received from the United reservations agent

	Poor Among the worst	Fair Not as Good as Most	Good About the Same as Most	Very Good Better than Most	Excellent Among the Best
Courtesy/Friendliness	○	○	○	○	○
Knowledge/helpfulness	○	○	○	○	○
Efficiency on completing transaction	○	○	○	○	○

The rating is monadic. United then asked respondents another question:

Comparative Example:
Please rate today's flight attendants compared to flight attendants on other airlines on each of the following items.

	Poor Among the worst	Fair Not as Good as Most	Good About the Same as Most	Very Good Better than Most	Excellent Among the Best
Courtesy/Friendliness	○	○	○	○	○
Assistance in cabin before departure	○	○	○	○	○
Responsiveness to your needs	○	○	○	○	○

Ratings are used very widely because they are easier and faster to administer and yield data that are amenable to being analyzed as if they are interval-scaled. But there is a risk of lack of differentiation among the scores when the particular attributes are worded positively or are positive constructs, such as values, and the respondents end-pile their ratings toward the positive end of the scale. Such lack of differentiation may potentially reduce the variance of the items being rated and reduce the ability to detect relationships with other variables.

McCarty and Shrum (2000) offer an alternative to simple rating. Respondents first picked their most and least important values (or attributes or factors), and then rated them. The remaining values were then rated. Their results indicate that, compared with a simple rating of values, the most-least procedure reduces the level of end-piling and increases the differentiation of values ratings, both in terms of dispersion and the number of different rating points used.

Strongly Disaprove or disagree	Disapprove or disagree	Undecided or neither agree nor disagree	Approve or agree	Strongly approve or agree
○	○	○	○	○

4. Each response is given a numerical weight (e.g., +2, +1, 0, −1, −2, or +1 to +5).
5. The individual's *total-attitude score* is represented by the algebraic summation of weights associated with the items checked. In the scoring process, weights are assigned such that the direction of attitude— favorable to unfavorable—is consistent over items. For example, if a + 2 were assigned to "strongly approve/agree" for favorable items, a + 2 should be assigned to "strongly disapprove/disagree" for unfavorable items.
6. On the basis of the results of the pretest, the analyst selects only those items that appear to discriminate well between high and low *total* scorers. This may be done by first finding the highest and lowest quartiles of subjects on the basis of *total* score. Then, the mean differences on each *specific* item are compared between these high and low groups (excluding the middle 50 percent of subjects).
7. The 20 to 25 items finally selected are those that have discriminated "best" (i.e., exhibited the greatest differences in mean values) between high versus low total scorers in the pretest.
8. Steps 3 through 5 are then repeated in the main study.

When analysis is completed, many researchers assume only ordinal properties regarding the placement of respondents along the continuum. Nonetheless, two respondents could have the same total score even though their response patterns to individual items were quite different. That is, the single (summated) score ignores the details of just which items were agreed with and which ones were not. Moreover, the total score is sensitive to how the respondent reacts to the descriptive intensity scale.

Often, a researcher will reverse the polarity of some items in the set (i.e., word items negatively) as a way to overcome the possibility of acquiescence bias (being overly agreeable). Having positively and negatively worded statements hopefully forces respondents with strong positive or negative attitudes to read carefully and use both ends of a scale. A researcher should reverse the polarity of some items, but may need to adjust the scoring, as appropriate. That is, a "strongly agree" response to a positive statement and a "strongly disagree" to a negative statement should be scored the same, and so forth.

Another approach to wording the summated scale adapts the statements into a set of non-directional questions, thereby alleviating the problems associated with mixed-wording scales (Wong, Rindfleisch, & Burroughs, 2003). As an illustration, a non-directional format for one item would be:

> "How much pleasure do you get from traveling? [Very little…A great deal]"

In contrast, the normal Likert format for this item is:

> "Traveling gives me a lot of pleasure [strongly agree, agree, neither agree nor disagree, disagree, strongly disagree]"

Some final comments are in order. When using this format, Likert (1967) stated that a key criterion for statement preparation and selection should be that all statements be expressions of desired behavior and not statements of fact. Because two persons with decidedly different attitudes may agree on fact, it is recognized that direction is the only meaningful measure

8. **Non-directed questions give respondents excessive latitude.**
What suggestions do you have for improving tomato juice? The question is about taste, but the respondent may offer suggestions about texture, the type of can or bottle, mixing juices, or something related to use as a mixer or in recipes.

9. **Forcing answers.** Respondents may not want, or may not be able to provide the information requested. Privacy is an important issue to most people. Questions about income, occupation, finances, family life, personal hygiene and beliefs (personal, political, religious) can be too intrusive and rejected by the respondent.

10. **Non-exhaustive listings.** Do you have all of the options covered? If you are unsure, conduct a pretest using the "Other (please specify) __________" option. Then revise the question making sure that you cover at least 90% of the respondent answers.

11. **Unbalanced listings.** Unbalanced scales may be appropriate for some situations and biased in others. When measuring alcohol consumption patterns, one study used a quantity scale that made the heavy drinker appear in the middle of the scale with the polar ends reflecting no consumption and an impossible amount to consume. However, we expect all hospitals to offer good care and may use a scale of excellent, very good, good, fair. We do not expect poor care.

12. **Double barreled questions.** What is the fastest and most convenient Internet service for you? The fastest is certainly not the most economical. The double barreled question should be split into two questions.

13. **Independent answers.** Make sure answers are independent. For example the question "Do you think basketball players as being independent agents or as employees of their team?" Some believe that yes, they are both.

14. **Long questions.** Multiple choice questions are the longest and most complex. Free text answers are the shortest and easiest to answer. When you Increase the length of questions and surveys, you decrease the chance of receiving a completed response.

15. **Questions on future intentions.** Yogi Berra (Famous New York Yankees Baseball Player) once said that making predictions is difficult, especially when they are about the future. Predictions are rarely accurate more than a few weeks or in some case months ahead.

VALIDITY AND RELIABILITY OF MEASUREMENT

The content of a measurement instrument includes a subject, theme, and topics that relate to the characteristics being measured. However the measuring instrument does not include all of the possible items that could have been included. When measuring complex psychological constructs such as perceptions, preferences, and motivations, hard questions must be asked to identify the items most relevant in solving the research problem:

1. Do the scales really measure what we are trying to measure?
2. Do subjects' responses remain stable over time?
3. If we have a variety of scaling procedures, are respondents consistent in their scoring over those scales that purport to be measuring the same thing?

By solving these problems, we establish the validity and reliability of scaling techniques.

Note that our focus is only on the general concepts and measures of validity and reliability that are used in cross-sectional studies. There is little documented research on issues of measure reliability and validity for time-series analysis.

The achievement of scale reliability is, of course, dependent on how consistent the characteristic being measured is from individual to individual (homogeneity over individuals) and how stable the characteristic remains over time. Just how reliable a scaling procedure turns out to be will depend on the dispersion of the characteristic in the population, the length of the testing procedure, and its internal consistency. Churchill and Peter (1984) concluded that rating scale estimates were largely determined by measuring characteristics such as number of items in a scale, type of scale, and number of scale points. They further concluded that sampling characteristics and measurement development processes had little impact.

In general, a measurement of the reliability of a scale (or measurement instrument) may be measured by one of three methods: test-retest, alternative forms, or internal consistency. The basics of reliability in a marketing context are reviewed by Peter (1979).

Test-Retest

The *test-retest method* examines the stability of response over repeated applications of the instrument. Do we achieve consistent results, assuming that the relevant characteristics of the subjects are stable over trials? One potential problem, of course, is that the first measurement may have an effect on the second one. Such effects can be reduced when there is a sufficient time interval between measurements. If at all possible, the researcher should allow a minimum of two weeks to elapse between measurements. Reliability may be estimated by any appropriate statistical technique for examining differences between measures.

Alternative Forms

The *alternative forms method* attempts to overcome the shortcomings of the test-retest method by successively administering equivalent forms of the measure to the same sample. Equivalent forms can be thought of as instruments built such that the same types and structures of questions are included on each form, but where the specific questions differ. The forms of the measurement device may be given one after the other or after a specified time interval, depending upon the investigator's interest in stability over time. Reliability is estimated by correlating the results of the two equivalent forms.

Internal Consistency

Internal consistency refers to estimates of reliability within single testing occasions. In a sense it is a modification of the alternative form approach, but differs in that alternatives are formed by grouping variables. The basic form of this method is split-half reliability, in which items are divided into equivalent groups (say, odd- versus even-numbered questions, or even a random split) and the item responses are correlated. In practice, any split can be made.

A potential problem arises for split-half in that results may vary depending on how the items are split in half. A way of overcoming this is to use coefficient alpha, known also as Cronbach's alpha, which is a type of mean reliability coefficient for all possible ways of splitting an item in half (Cronbach, 1951). Whenever possible, alpha should be used as a measure of the internal consistency of multi-item scales. Alpha is perhaps the most widely used measure of internal consistency for multiple-item measures within marketing research. One caution however, is that there should be a sufficient number of items in the measure so that alpha becomes meaningful. Alpha has been used for as few as two items, and this essentially amounts to a simple correlation between the two. Although there is no generally acceptable heuristic covering the number of items, common sense would indicate that the minimum number of items should be four or perhaps even six. What is clear, however, is that alpha is a function of the

Litwin, M. S. (2003). *How to assess and interpret survey psychometric* (2nd ed.). Thousand Oaks, CA: Sage.

McCarty, J. A., & Shrum, L. J. (2000). The measurement of personal values in survey research: A test of alternative rating procedures. *Public Opinion Quarterly, 64,* 271–298.

Morgan, M. (2003, March 31). Be careful with survey data. *Marketing News*, *37,* 26.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. Urbana: University of Illinois Press.

Peter, J. P., & Churchill, G. A., Jr. (1986, February). Relationships among research design choices and psychometric properties of rating scales: A meta-analysis. *Journal of Marketing Research*, *23*, 1–10.

Peterson, R. A. (2000). *Creating effective questionnaires*. Thousand Oaks, CA: Sage.

Payne, S. L. (1951). *The art of asking questions*. Princeton, NJ: Princeton University Press.

Rasinski, K. A. (1989). The effect of question wording on public support for government spending. *Public Opinion Quarterly*, *53*, 388–394.

Reuman, D. A. (1982). Ipsative behavioral variability and the quality of thematic apperceptive measurement of the achievement motive. *Journal of Personality and Social Psychology*, *43*(5), 1098–1110.

Semon, T. T. (1999, August 30). Scale ratings always betrayed by arithmetic. *Marketing News*, *33,* 7.

Semon, T. T. (2001, October 8). Symmetry shouldn't be goal for scales. *Marketing News*, *35,* 9.

Shimp, T. A., & Sharma, S. (1987, August). Consumer ethnocentrism: Construction and validation of the CETSCALE. *Journal of Marketing Research*, *24*, 280–289.

Spector, P. E. (1992). *Summated rating scale construction: An introduction*. Newbury Park, CA: Sage.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science, 103,* 677–680.

Sudman, S., & Bradburn, N. (1983). *Asking questions*. San Francisco: Jossey-Bass.

Thurstone, L. L (1959). *The measurement of values*. Chicago: University of Chicago Press.

Torgerson, W. S. (1958). *Theory and methods of scaling*. New York: Wiley.

Voss, K. E., Stem, D. E, Jr., and Fotopoulos, S. (2000). A comment on the relationship between coefficient alpha and scale characteristics. *Marketing Letters*, *11*(2), 177—191.

Whitlark, D., & Smith, S. (2004). "Pick and Choose " Marketing Research Volume 16, Issue 4, Pages 8-14, American Marketing Association, Chicago, December, 2004

Wong, N., Rindfleisch, A., & Burroughs, J. E. (2003, June). Do reverse-worded items confound measures in cross-cultural research? The case of the material values scale. *Journal of Consumer Research*, *30*, 72–91.

Chapter 11
HYPOTHESIS TESTING AND UNIVARIATE ANALYSIS

Scientific research is directed at the inquiry and testing of alternative explanations of what appears to be fact. For behavioral researchers, this scientific inquiry translates into a desire to ask questions about the nature of relationships that affect behavior within markets. It is the willingness to formulate hypotheses capable of being tested to determine (1) what relationships exist, and (2) when and where these relationships hold.

The first stage in the analysis process is identified to include editing, coding, and making initial counts of responses (tabulation and cross tabulation). In the current chapter, we then extend this first stage to include the testing of relationships, the formulation of hypotheses, and the making of inferences.

In *formulating hypotheses* the researcher uses "interesting" variables, and considers their relationships to each other, to find suggestions for working hypotheses that may or may not have been originally considered. In *making inferences,* conclusions are reached about the variables that are important, their parameters, their differences, and the relationships among them. A *parameter* is a summarizing property of a collectivity—such as a population—when that collectivity is not considered to be a sample (Mohr, 1990, p. 12).

Although the sequence of procedures, (a) formulating hypotheses, (b) making inferences and (c) estimating parameters is logical, in practice these steps tend to merge and do not always follow in order. For example, the initial results of the data analysis may suggest additional hypotheses that in turn require more and different sorting and analysis of the data. Similarly, not all of these steps are always required in a particular project; the study may be exploratory in nature, which means that it is designed more to formulate the hypotheses to be examined in a more extensive project, than to make inferences or estimate parameters.

AN OVERVIEW OF THE ANALYSIS PROCESS

The overall process of analyzing and making inferences from sample data can be viewed as a process of refinement that involves a number of separate and sequential steps that may be identified as part of three broad stages:

1. *Tabulation*: identifying appropriate categories for the information desired, sorting the data by categories, making the initial counts of responses, and using summarizing measures to provide economy of description and thereby facilitate understanding.
2. *Formulating additional hypotheses*: using the inductions derived from the data concerning the relevant variables, their parameters, their differences, and their relationships to suggest working hypotheses not originally considered.
3. *Making inferences*: reaching conclusions about the variables that are important, their parameters, their differences, and the relationships among them.

$$\text{Sample std. dev.} = s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{(1-1.5)^2 + (2-1.5)^2}{(2-1)}} = .71$$

$$\text{Est. std. error} = S_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{.71}{\sqrt{2}} = .5$$

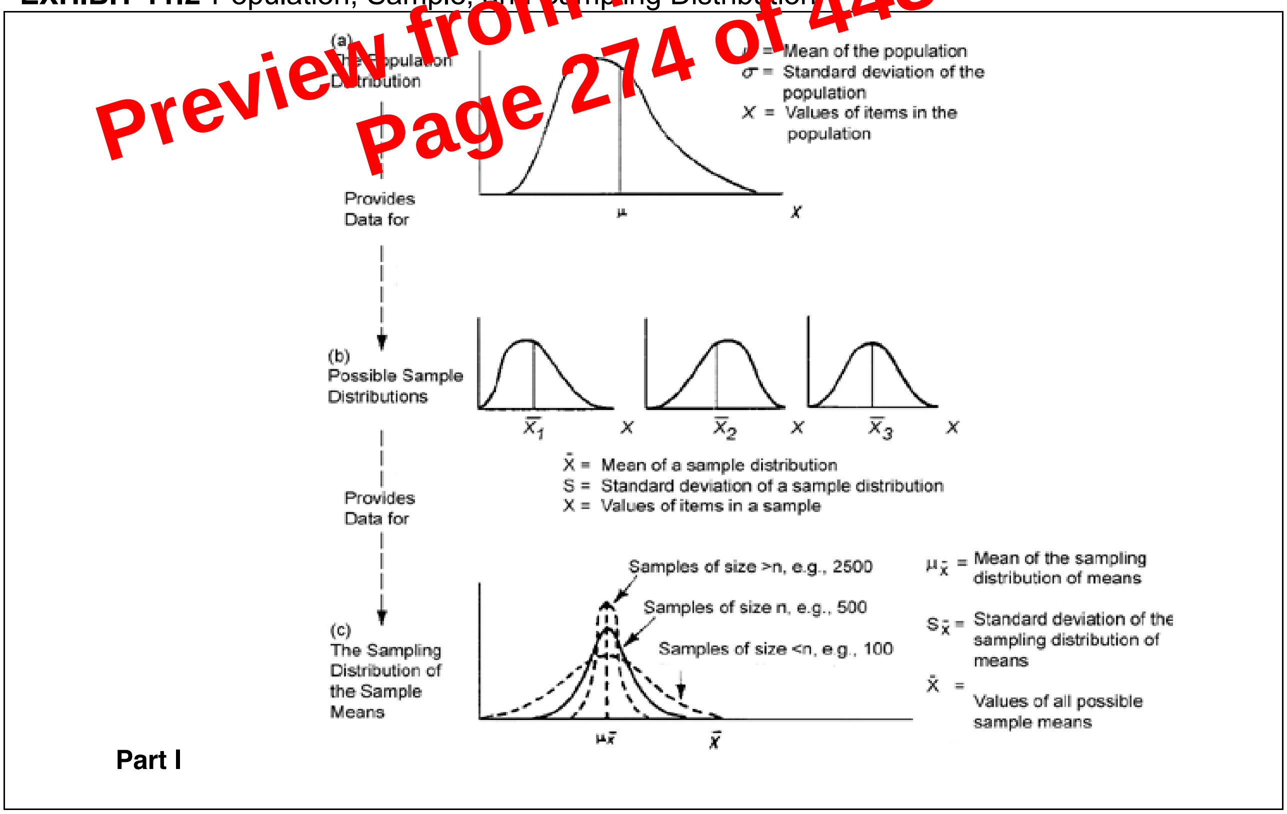
Given this $\bar{x}$ and $S_{\bar{X}}$, we can now estimate with a given probability, the intervals that give a range of possible values that could include μ, the population mean. For this single sample, they are:

68% = 1.5 ± 6.31 (.5) or -1.655 to 4.655
95% = 1.5 ± 12.71 (.5) or -4.855 to 7.855
99% = 1.5 ± 31.82 (.5) or -14.41 to 17.41

Thus, we could state that we are 99% confident that the population mean would fall within the interval –14.41 to 17.41. We note that this range is very wide and does include our population mean of 4.0. The size of the range is large because the small sample size (n=2). As the sample size increases, the numbers become larger, and gradually approximate a standard normal distribution.

The above discussion is summarized graphically in Exhibit 11.2. Part I of Exhibit 11.2 shows the relationship between the population, the sample, and the sampling distribution while Part II illustrates the impact of sample size on the shape of the sampling distribution for differently shaped population distributions.

EXHIBIT 11.2 Population, Sample, and Sampling Distribution



interval or ratio scaled data. In cases where data can be obtained only using ordinal or categorical scales the interpretation of the results may be questionable especially if the ordinal categories are not of equal interval. When data do not meet the rigorous assumptions of parametric method, we must rely on non-parametric methods which free us of the assumptions about the distribution.

Whereas parametric methods make inferences about parameters of the population (μ and s), non-parametric methods may be used to compare entire distributions that are based on nominal data. Other non-parametric methods that use an ordinal measurement scale test for the ordering of observations in the data set.

Problems that may be solved with parametric methods may often be solved by a non-parametric method designed to address a similar question. Often times, the researcher will find that the same conclusion regarding significance is made when data are analyzed by a parametric method and by its "corresponding" non-parametric method. We will now discuss a univariate parametric and non-parametric analyses. In the next chapter bi-variate parametric and non-parametric analyses are presented. Additional non-parametric analyses are presented in the appendix to Chapter 12.

Univariate Analyses of Parametric Data

Marketing researchers are often concerned with estimating parameters of a population. In addition, many studies go beyond estimation and compare population parameters by testing hypotheses about differences between them. Very often, the means, proportions and variances are the summary measures of concern. Our concern at this point is with differences between the means and proportions of the sample and those of the population as a whole. These comparisons involve a single variable. In the following sections, we will demonstrate three important concepts: (1) how to construct and interpret a confidence interval; (2) how to perform a hypothesis test; and (3) how to determine the power of a hypothesis test. These issues are discussed in more depth by Mohr (1990).

The Confidence Interval

The concept of a *confidence interval* is central to all parametric hypothesis testing. The confidence interval is *a range of values with a given probability (.95, .99, etc.) of covering the true population parameter*.

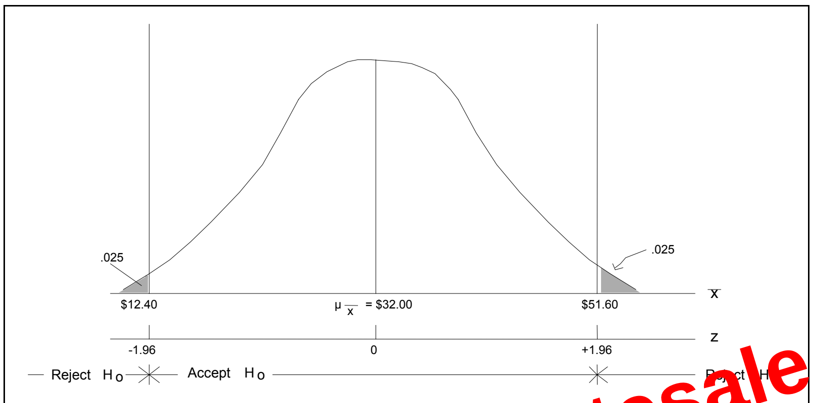
For example, assume we have a normally distributed population with population mean μ and a known population variance σ^2. Suppose we sample one item from the population, X. This single item is an estimate of μ, the population mean. Further, because the single item has been drawn randomly from a normally distributed population, the possible distribution of x values is the same as the population. This normal distribution permits us to estimate the probability associated with various intervals of values of X. For example, $p(-1.96 \leq z \geq +1.96) = .95$. Or, about 95% of the area under the normal probability curve is within this range. The confidence interval shows both the z values and the values that are included in the confidence interval. We compute this range for the sample problem that follows.

Suppose that it is a well-known fact that the average supermarket expenditure on laundry and paper products is normally distributed, with a mean of μ=\$32.00 per month, and the known population standard deviation is σ =10.00. The 95% confidence interval about the population mean is computed as:

$$\text{Lower limit} \leq \text{population mean} \geq \text{upper limit}$$
$$\mu - 1.96\sigma \leq \mu \geq \mu + 1.96\,\sigma$$
$$32 - (1.96)(10) \leq 32 \geq 32 + (1.96)(10)$$
$$32 - 19.6 \leq 32 \geq 32 + 19.6$$
$$12.4 \leq 32 \geq 51.6$$

Thus, we expect that 95% of all household expenditures will fall within this range, as shown in Figure 11.5.

Figure 11.5 Sampling Distribution of the Mean ($u_{\bar{x}}$)



If we expand this analysis to construct a confidence interval around the mean of a sample rather than the mean of a population, we must rely on a sampling distribution to define our normal distribution. The sampling distribution is defined by the means of all possible samples of size n. Recall that the population mean μ is also the mean of the normally distributed sampling distribution, whereas $\mu \pm z\sigma$ describes the confidence interval for a population, the value ($\pm\, \alpha\; s/\sqrt{n}$)) describes the confidence interval for the sampling distribution. This is the probability that this specified area around the sample mean covers the population mean. It is interesting to note that because n, the sample size, is included in the computation of the "standard error," we may estimate the population mean with any desired degree of precision, simply by having a large enough sample size.

Univariate Hypothesis Testing of Means

Where the Population Variance is Known

Researchers often desire to test a sample mean to determine if it is the same as the population mean. The z statistic describes probabilities of the normal distribution and is the appropriate tool to test the difference between μ, the mean of the sampling distribution, and, the sample mean when the population variance is known. The z statistic may, however, be used only when the following conditions are met:

(1) Individual items in the sample must be drawn in a random manner.
(2) The population must be normally distributed. If this is not the case, the sample must be large (>30), so that the sampling distribution is normally distributed.
(3) The data must be at least interval scaled.
(4) The variance of the population must be known.

1. When these conditions are met, or can at least be reasonably assumed to exist, the traditional hypothesis testing approach is as follows:

   (1) The null hypothesis (H_0) is specified that there is no difference between μ and $\bar{x}$. Any observed difference is due solely to sample variation.
   (2) The alpha risk (Type I error) is established (usually .05).
   (3) The z value is calculated by the appropriate z formula:

$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

   (4) The probability of the observed difference having occurred by chance is determined from a table of the normal distribution (Appendix B, Table B-1).
   (5) If the probability of the observed differences having occurred by chance is greater than the alpha used, then H_0 cannot be rejected and it is concluded that the sample mean is drawn from a sampling distribution of the population having mean μ.

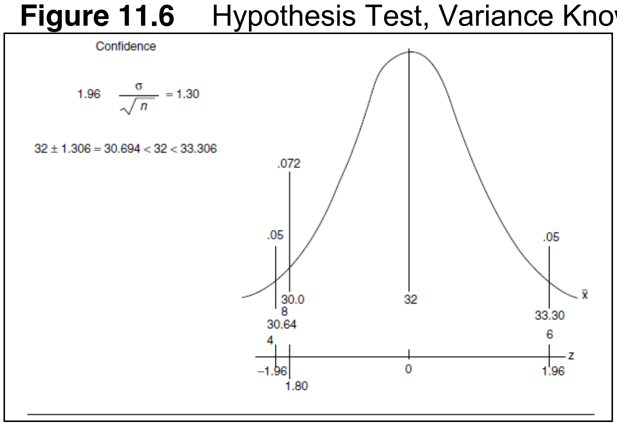
Returning to our supermarket example, suppose now that a random sample of 225 respondents was collected. The sample has a mean of =$30.80 and the population standard deviation is known to be $10.00. We want to know if the population mean μ=$32.00 equals the sample mean =30.80, given sample variation.

Steps:

1. All assumptions are met
2. $H_0 : \mu = \bar{x}$
3. $z = \dfrac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \dfrac{30.80 - 32.00}{10/\sqrt{225}} = -1.20/0.667 = -1.80$
4. $P(z = 1.80) = .928$
   $1 - P(z = 1.80) = .072$
5. Decision: Do not reject

The results are shown in Figure 11.6.

Figure 11.6 Hypothesis Test, Variance Known



Univariate Analysis: Test of a Proportion

The standard normal distribution may be used to test not only means, as explained above, but also differences in proportions. The univariate test of proportions, like the univariate test of means, compares the population proportion to the proportion observed in the sample. For a sample proportion, p,

$$Z = \frac{p - \pi}{S_p}$$

Where s_p, the estimated standard error of the proportion,

$$S_p = \sqrt{pq/n} = \sqrt{\frac{p(1-p)}{n}}$$

z = standard normal value
p = the sample proportion of successes
q = (1-p) = the sample proportion of failures
n = sample size

In a simple example, suppose the marketing manager of a snack food company is evaluating a new snack. 225 respondents are surveyed at a local shopping mall. The survey indicates that 87% are favorable toward the snack. The manager needs a 90% favorability rate. Is it safe to say that this is simply sampling variation?

$$Z = 1.338 = \frac{.87 - .90}{.022} \quad \text{where} \quad S_p = \sqrt{\frac{(.87)(.13)}{225}} = .022$$

given $\alpha = .05$ with $z = 1.96$, we cannot reject the null hypothesis.

CI = .90 - 1.96(.022) < 90 < 90 + 1.96(.022)

CI = .856 < .90 < .944

In this example the manager could (in a statistical sense) claim that 87% approval is no different that 90% approval with sampling variation. In terms of corporate policy, however, set cut points may be more rigidly held.

SUMMARY

Chapter 11 has introduced the basic concepts of formulating hypothesis testing and making statistical inference in the context of univariate analysis. In actual research, the analyst may alternate between analyzing the data and formulating hypotheses.

A hypothesis is a statement that variables (measured constructs) are related in a specific way. The null hypothesis, H_0, is a statement that no relationship exists between the variables tested or that there is no difference.

Statistics are based on making inferences from the sample of respondents to the population of all respondents by means of a sampling distribution. The sampling distribution is a distribution of the parameter values (means or variances) that are estimated when all possible samples are collected.

In Table 12.1, the control variable is the experimental area (test versus control) and the dependent variable is awareness. When comparing awareness in the test and control areas, row percentages are preferred. We note that before the spot TV campaign the percentage of respondents who are aware of Life is almost the same between test and control areas: 42 percent and 40 percent, respectively.

However, after the campaign the test-area awareness level moves up to 66 percent, whereas the control-area awareness (42 percent) stays almost the same. The small increase of 2 percentage points reflects either sampling variability or the effect of other factors that might be serving to increase awareness of Life in the control area.

On the other hand, computing percentages across the independent variable (column percent) makes little sense. We note that 61 percent of the aware group (before the spot TV campaign) originates from the test area; however, this is mainly a reflection of the differences in total sample sizes between test and control areas.

After the campaign we note that the percentage of aware respondents in the control area is only 33 percent, versus 39 percent before the campaign. This may be erroneously interpreted as indicating that spot TV in the test area depressed awareness in the control area. But we know this to be false from our earlier examination of raw percentages.

It is not always the case that one variable is clearly the independent or control variable and the other is the dependent or criterion variable. This should pose no particular problem as long as we agree, for analysis purposes, which variable is to be considered the control variable. Indeed, cases often arise in which each of the variables in turn serves as the independent and dependent variable.

Table 12.1 Awareness of Life Salad Dressing—Before and After Spot TV

	Before Spot TV				After Spot TV		
	Aware	Not Aware	Total	Area	Aware	Not Aware	Total Area
Test Area				**Test Area**			
Freq.	250	350	600	Freq.	330	170	550
Col %	61%	59%	60%	Col %	67%	44%	57%
Row %	42%	58%		Row %	66%	34%	
Control Area				**Control Area**			
Freq.	160	240	400	Freq.	160	220	380
Col %	39%	41%	40%	Col %	33%	56%	43%
Row %	40%	60%		Row %	42%	58%	
Total	410	590	1000	Total	490	390	880
Before TV Spot	41%	59%	100%	After TV Spot	56%	44%	100%

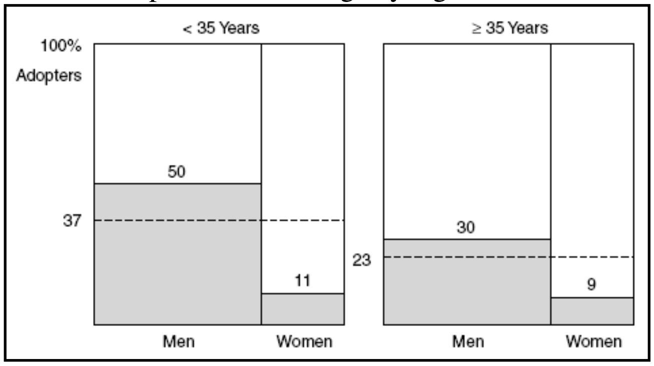
Interpretation of the Percentage Change

A second problem that arises in the use of percentages in cross-tabulations is the choice of which method to use in measuring differences in percentages. There are three principal ways to portray percentage change:

1. The absolute difference in percentages
2. The relative difference in percentages
3. The percentage of possible change in percentages

The same example can be used to illustrate the three methods.

Figure 12.1 Adoption—Percentage by Age and Gender



It is easy to see from the figure that adoption differs by age group (37 percent versus 23 percent). Furthermore, the size of the difference depends on the gender of the respondent: Men display a relatively higher rate of adoption, compared with women, in the younger age category.

Recapitulation

Representatives of three-variable association can involve many possibilities that could be illustrated by the preceding adoption-age-gender example:

1. In the example presented, adoption and age exhibit initial association; this association is still maintained in the aggregate but is refined by the introduction of the third variable, gender.

2. Adoption and age do not appear to be associated. However, adding and controlling on the third variable, gender, reveals suppressed association between the first two variables within the separate categories of men and women. In the two-variable cases, men and women exhibit opposite patterns, canceling each other out.

Although the preceding example was contrived to illustrate the concepts, the results are not unusual in practice. It goes almost without saying that the introduction of a third variable can often be useful in the interpretation of two-variable cross-tabulations.

However, the reader should be aware of the fact that we have deliberately used the phrase *associated with* rather than *caused by*. Association of two or more variables does not imply causation, and this statement is true regardless of our preceding efforts to refine some observed two-variable association through the introduction of a third variable.

In principle, of course, we could cross-tabulate four or even more variables with the possibility of obtaining further insight into lower-order (e.g., two-variable) associations. However, somewhere along the line, a problem arises in maintaining an adequate cell size for all categories. Unless sample sizes are extremely large in the aggregate and the number of categories per variable is relatively small, cross-tabulations rarely can deal with more than three variables at a time. A further problem, independent of sample size, concerns the high degree of complexity of interpretation that is introduced by cross-tabulations involving four or more variables. In practice, most routine applications of cross-tabulation involve only two variables at a time.

percent (87/320) of the 90 = 34.1 would be moderately loyal, and 28.3 percent (65/320) of the 90 = 25.4 would be brand switchers. In a similar fashion we can compute theoretical frequencies for each cell on the null hypothesis that loyalty status is statistically independent of occupational status. (It should be noted that the frequencies are the same, whether we start with the percentage of the row or the percentage of the column.)

The theoretical frequencies (under the null hypothesis) are computed and appear in parentheses in Table 12.5. The chi-square statistic is then calculated (and shown in the table) for each of the data cells in the table using the observed and theoretical frequencies:

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_i - F_i)^2}{F_i}$$

where f_i = actual observed frequency, F_i = theoretical expected frequency, and k = number of cells (r x c).

The appropriate number of degrees of freedom to use in this example is four. In general, if we have R rows and C columns, the degrees of freedom associated with the chi-square statistic are equal to the product

$$(R - 1)(C - 1)$$

If we use a significance level of 0.05, the probability table value of chi-square is 11.488. Hence, because the computed χ^2 of 21.08 is greater than the table value of 11.488, we reject the null hypothesis of independence between the characteristics loyalty status and occupational status.

A correction factor must be applied to the formula for chi-square when the number of observations in a cell is less than 10 (or where a 2 × 2 contingency table is involved). The numerator within the summation sign becomes $(|f_i - F_i| - 1/2)^2$ where the value 1/2 is the *Yates continuity correction*. This correction factor adjusts for the use of a continuous distribution to estimate probability in a discrete distribution.

Chi-square analysis of independence can be extended to deal with more than two variables. No new principles are involved. Three characteristics of the technique should be borne in mind, however. First, chi-square analysis deals with counts (frequencies) of data. If the data are expressed in percentage form, they should be converted to absolute frequencies. Second, the technique assumes that the observations are drawn independently. Third, the chi-square statistic cannot describe the relationship; it only gauges its statistical significance, regardless of logic or sense (Semon, 1999).

To assess the nature of the relationship, the researcher must look at the table and indicate how the variables appear to be related—a type of eyeball approach. This may involve examining any of the following combinations: (a) the variable combinations that produce large χ^2 values in the cells; (b) those with a large difference between the observed and expected frequencies; or (c) those where the cell frequency count, expressed as a percentage of the row total, is most different from the total column percentage (marginal column %). When variables are ordered or loosely ordered, a pattern can sometimes be observed by marking cells with higher than expected observed frequencies with a (+) and those with lower than expected observed frequencies with a (–), or even graphing the deviations from expected values, or cell X^2 values using a 3-dimensional contour map.

Exhibit 12.3 Look at Your Data Before You Analyze

In deciding which type of regression approach to use, it is important that the researcher know the *shape* of the interrelationship. The shape of the interrelationship is easy to see on a scatter diagram. Looking at this visually helps decide whether the relationship is, or approximates being, linear or whether it has some other shape which would require a transformation of the data—by converting to square roots or logarithms--or treatment as nonlinear regression (Semon, 1993).

Examination of the data by scatter diagrams also allows the researcher to see if there are any "outliers"—i.e., cases where the relationship is unusual or extreme as compared to the majority of the data points. A decision has to be made whether to retain such outliers in the data set for analysis.

When the regression line itself is included on the scatter diagram, comparisons between actual values and the values estimated by the regression formula can be compared and used to assess the estimating error. Of course, what the analyst is seeking is a regression function that has the *best fit* for the data, and this is typically based on minimizing the squares of the distances between the actual and estimated values—the so-called *least-squares* criterion.

The equation for a linear model can be written $\hat{Y} = a + bx$, where $\hat{Y}$ denotes values of the criterion that are predicted by the linear model; *a* denotes the intercept, or value of $\hat{Y}$ when X is zero; and *b* denotes the slope of the line, or change in $\hat{Y}$ per unit change in X.

But how do we find the numerical values of *a* and *b*? The method used in this chapter is known as *least squares* as discussed in Exhibit 12.3. As the reader will recall from introductory statistics, the method of least squares finds the line whose sum of squared differences between the observed values Y_i and their estimated counterparts $\hat{Y}_i$ (on the regression line) is a minimum.

Parameter Estimation

To compute the estimated parameters (*a* and *b*) of the linear model, we return to the data of Table 12.11. In the two-variable case, the formulas are relatively simple:

$$b = \frac{\sum YX_1 - n\overline{Y}\,\overline{X}}{\sum X^2 - n\overline{X}^2}$$

$$= \frac{247 - 10(4.3)(4.3)}{255 - 10(4.3)^2}$$

$$= 0.886$$

where *n* is the sample size and Y and X denote the mean of *Y* and *X*, respectively. Having found the slope *b*, the intercept *a* is found from

$$a = \overline{Y} - b\overline{X}$$
$$= 4.3 - 0.886(4.3)$$
$$= 0.491$$

leading to the linear function $\hat{Y} = 0.491 + 0.886X$

This function is drawn with the scatter plot of points in Figure 12.6. It appears to fit the plotted points rather well, and the model seems to be well specified (a linear rather than arc linear or other form seems to fit).

Assumptions of the Model

Underlying least-squares computations is a set of assumptions. Although least-squares regression models do not need to assume normality in the (conditional) distributions of the criterion variable, this assumption is made when we test the statistical significance of the contribution of the predictor variable in explaining the variance in the criterion (does it differ from zero?) With this in mind the assumptions of the regression model are as follows (the symbols α and β are used to denote population counterparts of *a* and *b*:

1. For each fixed value of X we assume a normal distribution of Y values exists. Our particular sample we assume that each y value is drawn independently of all others. What is being described is the "classical" regression model. Modern versions of the model permit the predictors to be random variables, but their distribution is not allowed to depend on the parameters of the regression equation.

2. The means of all of these normal distributions of Y lie on a straight line with slope β.

3. The normal distributions of Y all have equal variances. This (common) variance does not depend on values assumed by the variable X.

Because all values rarely fall on the regression line (only when the correlation is 1.0), there is unexplained error in predicting Y. This error is shown in Figure 12.3 as the difference between the values Y_i and the regression line $\hat{Y}$. For the population, our model is expressed algebraically as:

$$Y = \alpha + \beta X_1 + \varepsilon$$

where α = mean of Y population when $X_1 = 0$
β = change in Y population mean per unit change in X_1
ε = error term drawn independently from a normally distributed universe with mean μ(ε); the error term is independent of X_1

The nature of these assumptions is apparent in Figure 12.7. The reader should note that each value of X has associated with it a normal curve for Y (assumption 1). The means of all these normal distributions lie on the straight line shown in the figure (assumption 2).

However, like phi, the contingency coefficient is easy to compute from chi-square; moreover, like phi, its significance has already been tested in the course of running the chi-square test.

Both phi and the contingency coefficient are symmetric measures of association. Occasions often arise in the analysis of *R X C* tables (or the special case of 2 · 2 tables) where we desire to compute an *asymmetric* measure of the extent to which we can reduce errors in predicting categories of one variable from knowledge of the categories of some other variable. Goodman and Kruskal's *lambda-asymmetric coefficient* can be used for this purpose (Goodman & Kruskal, 1954).

To illustrate the lambda-asymmetric coefficient, let us return to the cross-tabulation of Table 12.15. Suppose that we wished to predict what category—no versus yes—a randomly selected person would fall in when asked the question, "Does your hair have enough body?" If we had no knowledge of the row variable (whether that person included "body" in her ideal set or not), we would have only the *column* marginal frequencies to rely on.

Our best bet, given no knowledge of the row variable, is always to predict "no," the *higher* of the column marginal frequencies. As a consequence, we shall be wrong in 41 of the 84 cases, a probability error of 41/84 = 0.49 Can we do better, in the sense of lower prediction errors, if we utilize information provided by the row variable?

If we know that "body" is included in the ideal set, we shall predict "no" and be wrong in only 8 cases. If we know that "body" is not included in the ideal set, we shall predict "yes" and be wrong in 17 cases. Therefore, we have reduced our number of prediction errors from 41 to 8 + 17 = 25, a decrease of 16 errors. We can consider this error reduction relatively:

$$\lambda_{C|R} = \frac{(\text{number of errors in first case}) - (\text{number of errors in second case})}{\text{numbers of errors in first case}}$$
$$= \frac{41 - 25}{41} = 0.39$$

In other words, 39 percent of the errors in predicting the column variable are eliminated by knowing the individual's row variable.

A less cumbersome (but also less transparent) formula for lambda-asymmetric is

$$\lambda_{C|R} = \frac{\sum_{k=1}^{K} f_{k_R^*} - F_c^*}{n - F_c^*} = \frac{(26 + 33) - 43}{84 - 43} = 0.39$$

Where f_{kr} *is* the *maximum* frequency found within each subclass of the row variable, F_c is the *maximum* frequency among the marginal totals of the column variable, and n is the total number of cases.

Lambda-asymmetric varies between zero, indicating no ability at all to eliminate errors in predicting the column variable on the basis of the row variable, and 1, indicating an ability to eliminate all errors in the column variable predictions, given knowledge of the row variable. Not surprisingly, we could reverse the role of criterion and predictor variables and find lambda-asymmetric for the row variable, given the column variable. In the case of Table 12.15, this results in λ=0.26.

Multicollinearity

Put simply, *multicollinearity* refers to the problem in applied regression studies in which the predictor variables exhibit very high correlation among themselves. This condition distorts the value of the estimated regression coefficients, inflates the standard error of beta, and thereby makes it more difficult to determine which predictor variable is having an effect.

Unless one is dealing with experimental design data, it is almost always the case that predictor variables in multiple regression will be correlated to some degree. The question is how much multicollinearity can be tolerated without seriously affecting the results? Unfortunately, there is no simple answer to this question.

The study of multicollinearity in data analysis revolves around two major problems: (a) How can it be detected; and (b) what can be done about it? These problems are particularly relevant to marketing research, where one often faces the dilemma of needing a large number of variables to achieve accuracy of predictors, and yet finds that as more predictors are added to the model, their intercorrelations become larger.

As indicated above, what constitutes serious multicollinearity is ambiguous. Some researchers have adopted various rules-of-thumb: For example, any pair of predictor variables must not correlate more than 0.9; if so, one of the predictors is discarded. While looking at simple correlations between pairs of predictors has merit, it can miss more subtle relationships involving three or more predictors. The above rule can be extended, of course, to the examination of *multiple* correlations between each predictor and all other predictors. Usually one would want to guard against having any of these multiple correlations exceed the multiple correlation of the criterion variable with the predictor set.

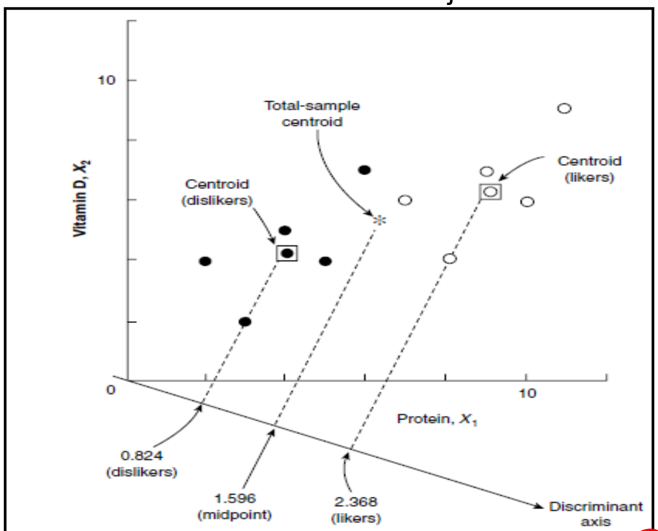
Essentially there are three procedures for dealing with multicollinearity:

1. Ignore it.
2. Delete one or more of the offending predictors.
3. Transform the set of predictor variables into a new set of predictor-variable combinations that are mutually uncorrelated.

Ignoring multicollinearity need not be as cavalier as it might sound. First, one can have multicollinearity in the predictor variables and still have strong enough effects that the estimating coefficients remain reasonably stable. Second, multicollinearity may be prominent in only a subset of the predictors, a subset that may not contribute much to accounted-for variance anyway. A prudent procedure in checking one's predictor set for multicollinearity is to examine the standard errors of the regression coefficients (which will tend to be large in the case of high multicollinearity). Second, one may randomly drop some subset of the cases (perhaps 20 percent or so), rerun the regression, and then check to see if the signs and relative sizes of the regression coefficients are stable. Third, most computer regression routines incorporate checks for serious multicollinearity (see SPSS Linear Regression Analysis Statistics Help for a list of diagnostic analyses); if the program does not indicate this condition, the researcher can generally assume that the problem is not acute.

If multicollinearity is severe, one rather simple procedure is to drop one or more predictor variables that represent the major offenders. Usually, because of their high intercorrelations with the retained predictors, the overall fit will not change markedly. Pragmatically, if a particular pair of predictors is highly collinear, one would retain that member of the pair whose discriminant axis as well.) We note that the discriminant axis favors X_1 (as we guessed it would) by giving about 2.5 times the (absolute-value) weight ($d_1 = 0.368$ versus $d_2 = -0.147$ to X_1) as is given to X_2.

Figure 13.6 Plot of the Discriminant Axis and Point Projections



The analysis output is shown in Exhibit 13.5. Note that the discriminant function coefficients are different from those shown earlier, meaning that the discriminant scores for each person also will differ. This is because the discriminant axis is *not run through the origin*, but rather there is an intercept term. The unstandardized discriminant function coefficients evaluated at group means sum to zero. The analysis is not affected by this, and all conclusions to be made will not differ whether or not the axis is through the origin.

EXHIBIT 13.5 Discriminant Analysis for Cereal Evaluations

Group Statistics

EVALUATE		Mean	Std. Deviation	Valid *N* (listwise) Unweighted	Weighted
GRP 1	PROTEIN	4.0000	1.58114	5	5.000
	VITAMIN	4.4000	1.81659	5	5.000
GRP 2	PROTEIN	9.0000	1.58114	5	5.000
	VITAMIN	6.4000	1.81759	5	5.000
Total	PROTEIN	6.5000	3.02765	10	10.000
	VITAMIN	5.4000	2.01108	10	10.000

Tests of Equality of Group Means

	Wilks' Lambda	F	df1	df2	Sig.
PROTEIN	.242	25.000	1	8	.001
VITAMIN	.725	3.030	1	8	.120

Summary of Canonical Discriminant Functions

<u>Eigenvalues</u>

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	3.860[a]	100.0	100.0	.891

First 1 canonical discriminant functions were used in the analysis.

Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-Square	df	Sig.
1	.206	11.068	2	.004

Classifying the Persons

It is well and good to find the discriminant function, but we are still interested in how well the function classifies the ten cases. The classification problem, in turn, involves two additional questions: (a) how well does the function assign the known cases in the sample; and (b) how well does it assign new cases *not* used in computing the function in the first place?

To answer these questions we need an *assignment rule*. One rule that seems intuitively plausible is based on Figure 13.6. A *classification boundary* between the two groups, Z_{crit}, can be identified as being midway between the means of the function for each of the two groups. To classify an individual, if $Z_i > Z_{crit}$ then the individual belongs in one group, while if $Z_i < Z_{crit}$ then the individual goes into the other group. As can be seen from Figure 13.6, no misassignments will be made if we adopt the rule:

- Assign all cases with discriminant scores that are on the left of the midpoint (1.596) to the *disliker* group.
- Assign all cases with discriminant scores that are on the right of the midpoint (1.596) to the *liker* group.

That is, all true dislikers will be correctly classified as such, as will all true likers. This can be shown by a $2 \cdot 2$ table, known as a classification or *confusion matrix*, as shown later in Table 13.6. We see that all entries fall along the main diagonal. For example, had any of the five true dislikers been called likers, the first row and second column would contain not a zero but, rather, the number of such misassignments.

The application of this rule can be stated in equivalent terms:

- Substitute the centroid of each group in the discriminant function and find the respective group scores (in our case, 0.824 for dislikers and 2.368 for likers).
- For any new case, compute the discriminant score and assign the case to that group whose group score is closer.

This rule makes two specific assumptions: (a) the prior probability of a new case falling into each of the groups is equal across groups; and (b) the cost of misclassification is equal across groups.

If a higher probability existed for likers, we could reduce the expected probability of misclassification by moving the cutting point of 1.596 to the left (closer to 0.824, the mean score for dislikers) so as to give a wider interval for the larger (likers) group. Similarly, if the cost of misclassifying a liker is higher than that for misclassifying a disliker, the cutting point would also be moved closer to 0.824.

A second point of interest concerns the tendency for classification tables (i.e., the confusion matrix) based on the sample used to develop the classification function to show better results than would be found upon cross-validation with new cases. That is, some capitalization on chance takes place in discriminant analysis, and one needs a way to measure this bias of an inflated percentage of correctly classified observations, just as cross-validation should be employed in multiple regression.

Procedures are available to develop a truer summary of the degree of correct assignments than would be obtained from fresh data (Lachenbruch, 1975; Klecka, 1980; Frank, Massy, & Morrison, 1965).

The most frequently suggested validation approach is the *holdout* method, in which the data set is randomly split into two subsamples. One subsample is used to develop the discriminant function and the other subsample is used as "fresh data" to test the function. It is suggested that this split-sample validation be replicated by using a different convention for the random assignment of the observations. Whether a researcher uses only one validation sample or replications, there may be problems associated with the sampling procedures used to arrive at the analysis and validation subsamples. These problems relate to the number of variables included in the discriminant function (when a stepwise procedure is used) and to inferences that the researcher might make concerning the classificatory power of resulting discriminant functions.

There may be differing numbers of variables included on any given replication. Also, there can be a wide range in the correct classification performance of the validation replications, although the analysis functions are quite stable.

There is another dimension to the validation process discussed above. For the analysis and validation samples there is a measure of the percentage of cases that would have been correctly classified based on *chance* alone. Obviously, the researcher is interested in whether the discriminant function is a significant improvement over chance classification.

Because our "Cereal Data" achieved 100 percent correct classification, we will now switch to another example to demonstrate these measures of chance classification. The classification or "confusion" matrix shown at the end of Table 13.6 was derived from a discriminant analysis where accident insurance purchasers and non-purchasers were predicted using demographic variables.

Overall, 74.16 percent of all subjects surveyed were correctly classified as purchasers and non-purchasers of accident insurance. This observed classification was found to be significant at the .001 level ($\chi 2$ = 24.59, df = 1) and (Q = 69.58, df = 1). Thus, observed classification is significantly different from expected chance classification.

Tests of Group Differences

Morrison (1969) considered the question of how well variables discriminate by formulating a likelihood ratio to estimate chance classification. This likelihood analysis provides a criterion that may be used to compare the proportion of correctly classified observations with the proportion expected by chance. This proportion, designated the proportional chance criteria, or Cpro (Morrison, 1969), is expressed as where

α = the proportion of customers in the sample categorized as purchasers

p = the true proportion of purchasers in the sample

$(1 - \alpha)$ = the proportion of the sample classified as non-purchasers

$(1 - p)$ = the true proportion of non-purchasers in the sample

This likelihood analysis states that 65.94 percent of the overall sample is expected to receive correct classification by chance alone. The proportional chance criterion, C_{pro}, has been used mainly as a point of reference for subjective evaluation (Morrison, 1969), rather than the basis of a statistical test to determine if the expected proportion differs from the observed proportion that is correctly classified.

$$R_j = \frac{I_j}{\sum_{j=1}^{n} I_j}$$

The end result of using this procedure is shown in Table 13.7, which is taken from a study of teenage smoking designed to see if consumption values (as represented by underlying had smoked and those who had never smoked (Albaum, Baker, Hozier, & Rogers, 2002). As these data show, although nine variables were included in the discriminant function, slightly more than 88 percent of the total discrimination was accounted for by only three variables.

Table 13.7 Relative Importance of Consumption Values for Teenage Smokers and Nonsmokers

Variables (Factors)	Standardized Coefficients	Unstandardized Coefficients (k_j)	Have Smoked Mean (X_1)	Never Smoked Mean (X_2)	$[k_j(X_{j1}-X_{j2})]$ (I_j)	Relative Importance Weight (R_j)
1	-0.298	-0.299	-0.0989	0.0908	0.0567	7.9%
2	0.091	0.091	0.0301	-0.0277	0.0053	0.7%
3	0.832	0.864	0.2861	-0.2627	0.4741	66.4%
4	0.344	0.345	0.1144	-0.1050	0.0757	10.6%
5	0.364	0.366	0.1211	-0.1112	0.0850	11.9%
6	-0.018	-0.018	-0.0061	0.0056	0.0002	0.0%
7	-0.004	-0,004	-0.0014	0.0013	0.0000	0.0%
8	-0.007	-0.007	-0.0023	0.0021	0.0000	0.0%
9	0.161	0.161	0.0533	[illegible]	0.0165	2.3%
					[illegible]	100.0%

Correctly Classified 60.2% (cross-validated grouped cases), Cpro – 51.8%
Wilk's Lambda - .886 (prob. < .01)
Canonical Correlation - .337
Source: Albaum, Baker, Hozier and Rogers, 2002, p.68.

Determining relative importance of the predictor variables in discriminant analysis becomes increasingly difficult when more than two groups are involved. Although various coefficients and indices can be determined, interpretation becomes critical since more than one discriminant function may be involved.

Multiple Discriminant Analysis

All of the preceding discussion regarding objectives and assumption structure applies to multiple discriminant analysis as well. Accordingly, discussion of this section will be comparatively brief. What primarily distinguishes *multiple discriminant analysis* from the two-group case is that *more than one* discriminant function may be computed. For example, if we have three groups we can compute, in general, two nonredundant discriminant functions (as long as we also have at least two predictor variables). In general, with G groups and m predictors we can find up to the lesser of $G - 1$, or m, discriminant functions.

Not all the discriminant functions may be statistically significant, however. Moreover, it turns out to be a characteristic of multiple discriminant analysis that the first function accounts for the highest proportion of the among-to within-groups variability; the second function, the next highest; and so on. Accordingly, we may want to consider only the first few functions, particularly when the input data are rather noisy or unreliable to begin with. There remains the problem of interpretation of the functions.

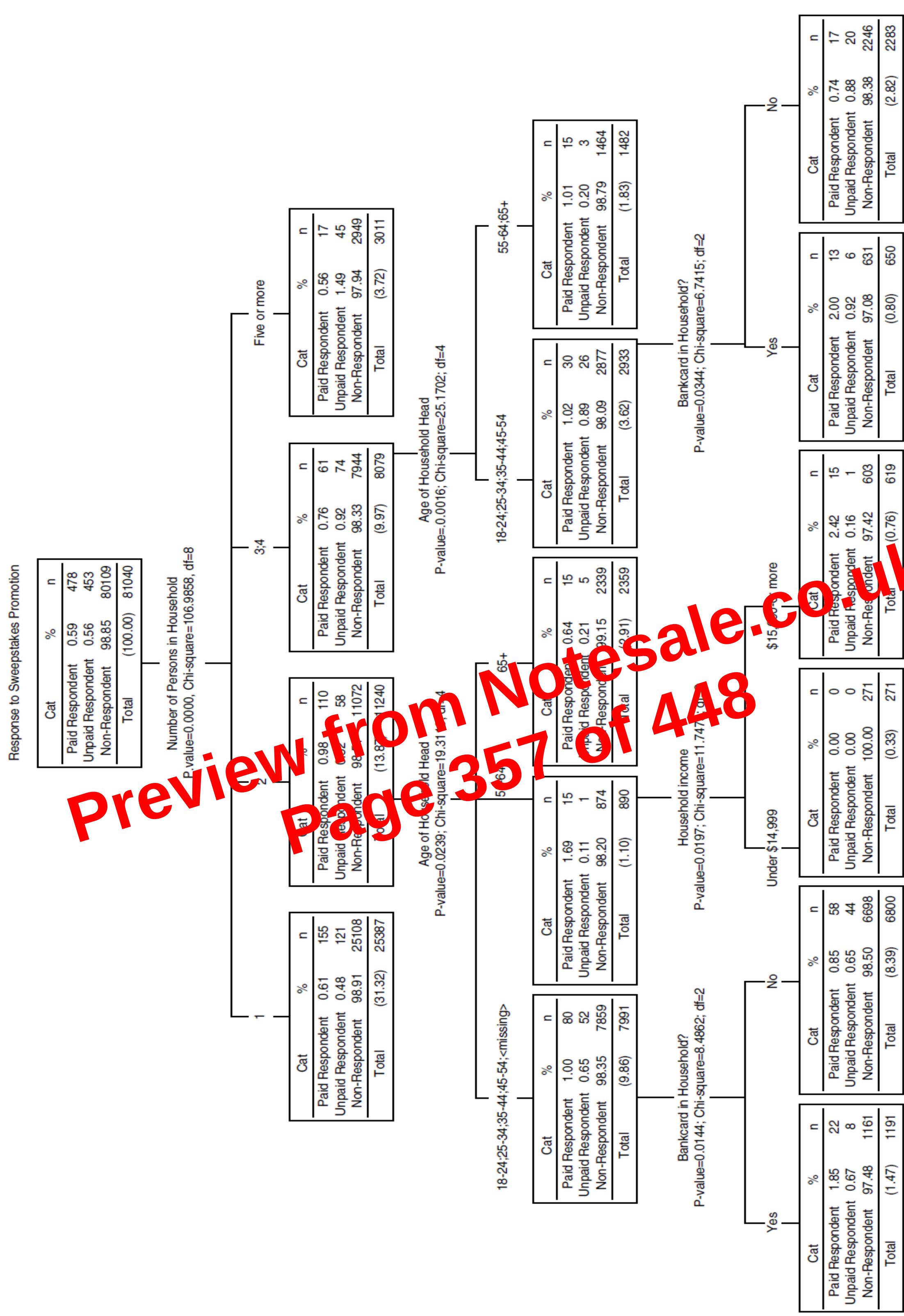


Figure 13.7 CHAID Sweepstakes Tree

Figure 13.8 CHAID Tree Summary Map

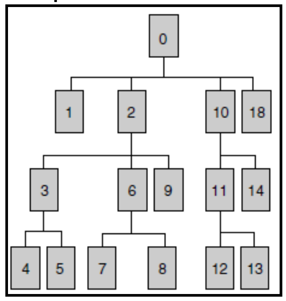


Table 13.8 CHAID Gain

Gain Summary								
Target variable: Response to Sweepstakes Promotion								
	Node-by-Node				*Cumulative*			
Node	*Node: n*	*Node:%*	*Gain*	*Index (%)*	*Node: n*	*Node:%*	*Gain*	*Index (%)*
8	619	0.76	0.69	3628.50972	619	0.76	0.69	3628.50972
12	650	0.80	0.49	2572.90367	1269	1.57	0.59	3087.81316
4	1191	1.47	0/45	2381.18629	2460	3.04	0.52	2745.70235
14	1482	1.83	0.19	1008.12275	3942	4.86	0.40	2092.45705
5	6800	8.39	0.11	554.14567	10742	13.26	0.21	1118.66098
9	2359	2.91	0.06	309.87554	13101	16.17	0.19	97[illegible].02898
13	2283	2.82	0.05	271.75404	15384	18.98	0.17	868.95912
1	25384	31.32	0.03	168.00454	40768	50.31	0.08	432.51311
18	3011	3.72	-0.05	-283.25238	79938	98.64	0.02	115.27446
7	271	0.33	-0.15	-787.99468	80209	98.97	0.02	122.22261

Node 12 is a 3–4 person household with a young to middle-age head of household who holds a bankcard.

The process of identifying profitable segments continues until we observe that Nodes 18 and 7 offer a negative gain (meaning they are unprofitable). We conclude that we should not target the large family segment having five or more family members (Node 18), or the poor empty-nester segment that is identical to our most profitable segment (Node 8), except without the income base.

The upshot of all of this is that by data mining with CHAID, we can identify 50.31 percent of our database that promises to be profitable and another 49.69 percent (39,441 individuals) that may be unprofitable. We further observe that the average gain (profit) drops from 0.17 to 0.08 with the inclusion of Node 1, which has an average gain of 0.03 per person for that segment. The successive targeting of the next most profitable segment can do much to maximize profits, as can eliminating the negative gain segments. Simply refining the database by deleting nonconforming profiles often does much to improve profitability of the marketing effort.

In short, this kind of information underscores the importance of certain types of demographic variables in this case, and leads to a more detailed understanding of the kinds of segments that responded by subscribing to this sweepstakes magazine offer.

While other summary statistics (e.g., probabilities, chi-square) can also be shown, the above outputs represent the principal ones. It should be emphasized that a basic assumption underlying the application of CHAID is that variables useful in explaining one part of the database are not necessarily those most effective for explaining another part.

Johnson, R. A., & Wichen, D. W. (2002). *Applied multivariate statistical analysis* (5th ed.). Upper Saddle River, NJ: Prentice Hall.

Klecka,W. R. (1980). *Discriminant analysis*. Beverly Hills, CA: Sage.

Lachenbruch, P. A. (1975). *Discriminant analysis*. New York: Hafner Press.

Levine, M. S. (1977). *Canonical analysis and factor comparison*. Beverly Hills, CA: Sage.

Liao, T. F. (1994). *Interpreting probability models: Logit, probit, and other generalized linear models*. Thousand Oaks, CA: Sage.

Magidson, J., & SPSS, Inc. (1993). *SPSS for Windows CHAID Release 6.0*. Chicago: SPSS, Inc. Malhotra, N. K. (1984, February). The use of linear logit models in marketing. *Journal of Marketing*

Research, 21, 20–31. Menard, S. (2001). *Applied logistic regression analysis* (2nd ed.). Thousand Oaks, CA: Sage.

Monroe, K. B., & Petroshius, S. M. (n.d.). *Developing causal priorities in marketing*. Working Paper, Virginia Polytechnic Institute and State University, Blacksburg, VA.

Morrison, D. G. (1969, May). On the interpretation of discriminant analysis. *Journal of Marketing Research, 6,* 156–163.

Mosteller, F., & Wallace, D. L. (1963, June). Influence in an authorship problem. *Journal of the American Statistical Association, 58,* 275–309.

Smith, S. M. (2004). *A note on the interpretation and analysis of the linear discriminant model for prediction and classification*. Retrieved May 18, 2004, from http://marketing.byu.edu/htmlpages/ tutorials/discriminant.htm

Steckel, J. H., & Vanhonacker, W. R. (1991). Crossvalidating regression models in marketing research. (Working Paper # MARK-91-10.) Stern School of Business, New York University.

Stevens, J. (1996). *Applied multivariate statistics for the social sciences*. Mahwah, NJ: Lawrence Erlbaum.

Thompson, B. (1984). *Canonical correlation analysis: Uses and interpretations*. Beverly Hills, CA: Sage.

Whitlark, D., & Smith, S. (2001, Summer). Using correspondence analysis to map relationships: It's time to think strategically about positioning and image data. *Marketing Research, 2,* 23–27. Whitlark, D., & Smith, S. (2004). *Measuring brand performance using online questionnaires: Advantages and issues with "pick any" data*. Provo, UT: Marriott School, Brigham Young Univ.

We can illustrate cluster analysis by a simple example. The problem is to group a set of twelve branches of a bank into three clusters of four branches each. Groups will be formed based on two variables, the number of men who have borrowed money ($X1$) and the number of women who have borrowed money ($X2$). The branches are plotted in two dimensions in the figure.

We use a proximity measure, based on Euclidean distances between any two branches. Branches 2 and 10 appear to be the closest together. The first cluster is formed by finding the midpoint between branches 2 and 10 and computing the distance of each branch from this midpoint (this is known as applying the *nearest-neighbor algorithm*). The two closest branches (6 and 8) are then added to give the desired-size cluster. The other clusters are formed in a similar manner. When more than two dimensions (that is, characteristics) are involved, the algorithms become more complex and a computer program must be used for measuring distances and for performing the clustering process.

Selecting the Clustering Methods

Once the analyst has settled on a pairwise measure of profile similarity, some type of computational routine must be used to cluster the profiles. A large variety of such computer programs already exist, and more are being developed as interest in this field increases. Each clustering program tends to maintain a certain individuality, although some common characteristics can be drawn out. The following categories of clustering methods are based, in part, on the classification of Ball and Hall (1964):

1. *Dimensionalizing the association matrix*. These approaches use principal-components or other factor-analytic methods to find a dimensional representation of points from *interobject* association measures. Clusters are then developed on the basis of grouping objects according to their pattern of component scores.

2. *Nonhierarchical methods*. The methods start right from the proximity matrix and can be characterized in three ways:

   a. *Sequential threshold*. In this case a cluster center is selected and all objects within a prespecified distance threshold value are grouped. Then a new cluster center is selected and the process is repeated for the unclustered points, and so on. (Once points enter a cluster, they are removed from further processing.)

   b. *Parallel threshold*. This method is similar to the preceding method, except that several cluster centers are selected simultaneously and points within a distance threshold level are assigned to the nearest center; thresholds can then be adjusted to admit fewer or more points to clusters.

   c. *Optimizing partitioning*. This method modifies categories (a) or (b) in that points can later be reassigned to clusters on the basis of optimizing some overall criterion measure, such as average within-cluster distance for a given number of clusters.

3. *Hierarchical methods*. These procedures are characterized by the construction of a hierarchy or tree-like structure. In some methods each point starts out as a unit (single point) cluster. At the next level the two closest points are placed in a cluster. At the following level a third point joins the first two, or else a second two-point cluster is formed based on various criterion functions for assignment. Eventually all points are grouped into one larger cluster. Variations on this procedure involve the development of a hierarchy from the top down. At the beginning the points are partitioned into two subsets based on some criterion measure related to average within-cluster distance. The subset with the highest average within-cluster distance is next partitioned into two subsets, and so on, until all points eventually become unit clusters.

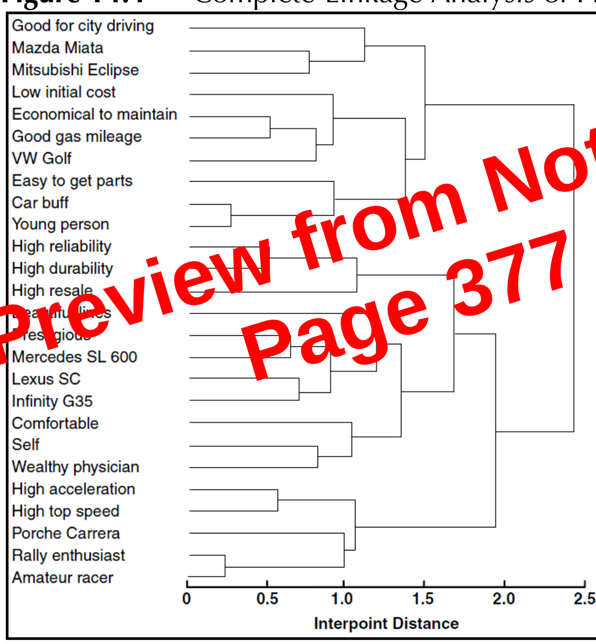
While the above classes of programs are not exhaustive of the field, most of the more widely used clustering routines can be classified as falling into one (or a combination) of the above categories. Criteria for grouping include such measures as average within-cluster distance and threshold cutoff values. The fact remains, however, that even the optimizing approaches achieve only conditional optima, since an unsettled question in this field is *how many* clusters to form in the first place.

A Product-Positioning Example of Cluster Analysis

Cluster analysis can be used in a variety of marketing research applications. For example, companies are often interested in determining how their products are positioned in terms of competitive offerings and consumers' views about the types of people most likely to own the product.

For illustrative purposes, Figure 14.4 shows the result of a hypothetical study conducted for seven sport cars, six types of stereotyped owners, and 13 attributes often used to describe cars.

Figure 14.4 Complete-Linkage Analysis of Product-Positioning Data



The inter-object distance data were based on respondents' degree-of-belief ratings about which attributes and owner types described each car. In this case, a complete-linkage algorithm was also used to cluster the objects (Johnson, 1967). The complete linkage algorithm starts by finding the two points with the minimum Euclidean distance. However, joining points to clusters is accomplished by maximizing the distance from a point in the first cluster to a point in the second cluster. Looking first at the four large clusters, we note the *car* groupings:

- Mazda Miata, Mitsubishi Eclipse
- VW Golf
- Mercedes 600 SL, Lexus SC, Infinity G35
- Porsche Carrera

Euclidean distance measure. A total of five clusters was derived based on the distance between countries and the centers of the clusters across the five predictor variables. The number of clusters selected was based on the criteria of total within-cluster distance and interpretability. A smaller number of clusters led to a substantial increase of within-cluster variability, while an increase in the number of clusters resulted in group splitting with a minimal reduction in distance. The composition of the clusters is shown in Table 14.4.

Computer Analyses

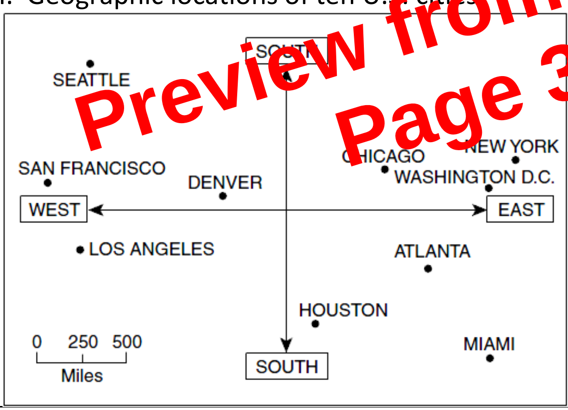
There are many computer programs available for conducting cluster analysis. Most analysis packages have one or more routines. Smaller, more specialized packages (such as PC-MDS) for cluster routines are also available. Finally, some academicians have developed their own cluster routines that they typically make available to other academicians for no charge.

MULTIDIMENSIONAL SCALING (MDS) ANALYSIS

Multidimensional scaling is concerned with portraying psychological relations among stimuli—either empirically-obtained similarities, preferences, or other kinds of matching or ordering—as geometric relationships among points in a multidimensional space. In this approach one represents *psychological dissimilarity as geometric distance*. The axes of the geometric space, or some transformation of them, are often (but not necessarily) assumed to represent the psychological bases or attributes along which the judge compares stimuli (represented as points or vectors in his or her psychological space).

Figure 14.5 Nonmetric MDS of 10 U.S. Cities

I. Geographic locations of ten U.S. cities



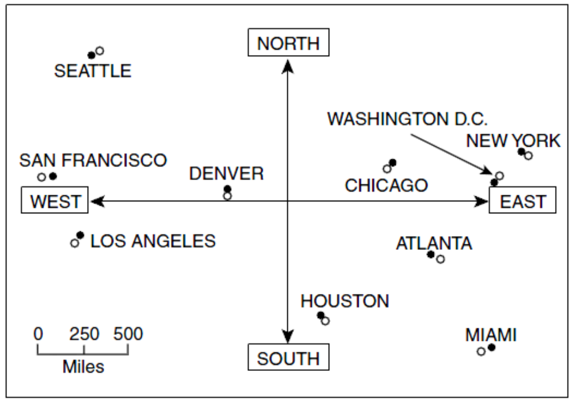
II. Airline distance between ten U.S. cities

CITIES	ATLA.	CHIC.	DENV.	HOUS.	L.A.	MIAMI	N.Y.	S.F.	SEAT.	WASH. D.C.
ATLANTA		587	1212	701	1936	604	748	2139	2182	543
CHICAGO			920	940	1745	1188	713	1858	1737	597
DENVER				879	831	1726	1631	949	1021	1494
HOUSTON					1374	968	1420	1645	1891	1220
LOS ANGELES						2339	2451	347	959	2300
MIAMI							1092	2594	2734	923
NEW YORK								2571	2408	205
SAN FRANCISCO									678	2442
SEATTLE										2329
WASHINGTON D.C.										

III. Nonmetric (Rank Order) Distance Data

CITIES	ATL	CHI	DEN	HOU	LA	MIA	NY	SF	SEA	D.C.
ATL		4	22	8	34	6	10	35	36	3
CHI			13	15	31	21	9	32	30	5
DEN				12	11	29	27	16	19	26
HOU					24	18	25	28	33	23
LA						39	42	2	17	37
MIA							20	44	45	14
NY								43	40	1
SF									7	41
SEA										38
D.C.										

IV. Original (•) and recovered (o) city locations via nonmetric MDS



preferred by that respondent. Although the original input data may be based on between-set relationships, if the simple unfolding model holds, one can also infer respondent to- respondent similarities in terms of the closeness of their ideal points to each other. Brand to- brand similarities may be analogously inferred, based on the relative closeness of pairs of brand points.

The point-vector model of two-mode, two-way data is a *projection* model in which one obtains respondent *i*'s preference scale by projecting the *J* brand points onto respondent *i*'s vector (Figure 14.5). Point-vector models also show ideal points or points of "ideal preference". This ideal point is located at the terminus or end of the vector. Projections are made by drawing a line so that it intersects the vector at a 90-degree angle. The farther out (toward vector *i*'s terminus) the *projection* is, the more preferred the brand is for that respondent.

Collecting Data for MDS

The content side of MDS—dimension interpretation, relating physical changes in products to psychological changes in perceptual maps—poses the most difficult problems for researchers. However, methodologists are developing MDS models that provide more flexibility than a straight dimensional application. For example, recent models have coupled the ideas of cluster analysis and MDS into hybrid models of categorical-dimensional structure. Furthermore, conjoint analysis, to be discussed next, offers high promise for relating changes in the physical (or otherwise controlled) aspects of products to changes in their psychological imagery and evaluation. Typically, conjoint analysis deals with preference (and other dominance-type) judgments rather than similarities. However, more recent research has extended the methodology to similarities judgments.

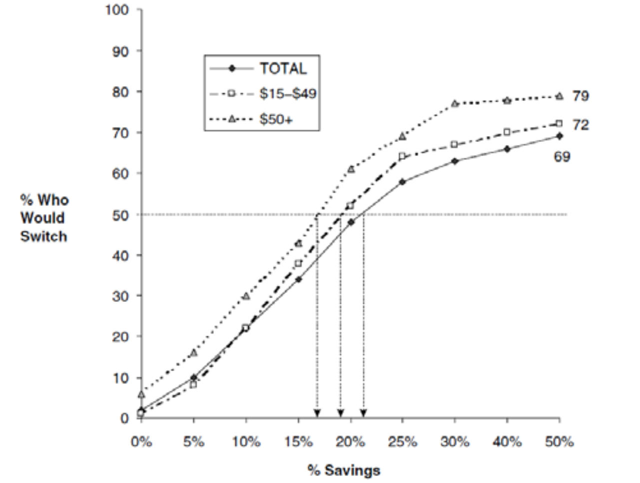
On the input side, there are issues that arise concerning data collection methods. In MDS studies, there are four most commonly used methods of data collection:

- *Sorting task*. Subjects are asked to sort the stimuli into a number of groups, according to similarity. The number of groups is determined by the subject during the judgment task.
- *Paired comparison task*. Stimuli are presented to subjects in all possible pairs of two stimuli. Each subject has to rate each pair on an ordinal scale (the number of points can vary) where the extreme values of the scale represent maximum dissimilarity and maximum similarity.
- *Conditional ranking task*. Subjects order stimuli on the basis of their similarity with an anchor stimulus. Each of the stimuli is in turn presented as the anchor.
- *Triadic combinations task*. Subjects indicate which two stimuli of combinations of three stimuli form the most similar pair, and which two stimuli form the least similar pair.

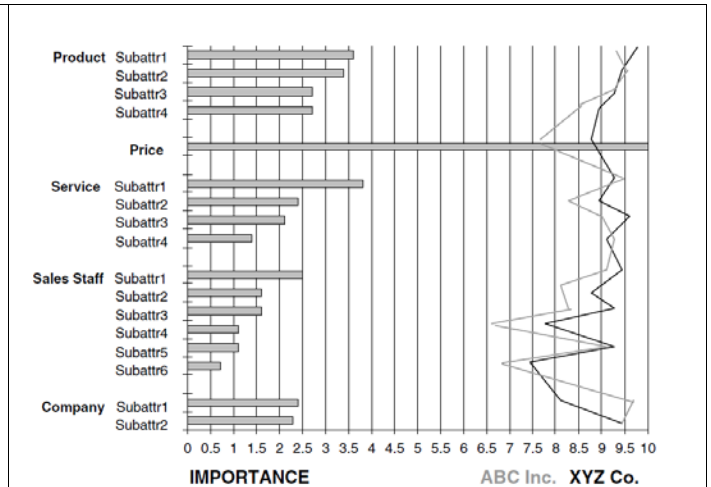
When subjects perform a similarity (or dissimilarity) judgment they may experience increases in fatigue and boredom (Bijmolt and Wedel,1995, p. 364).

Bijmolt and Wedel examined the effect of the alternative data collection methods on fatigue, boredom and other mental conditions. They showed that when collecting data, conditional rankings and triadic combinations should be used only if the stimulus set is relatively small, and in situations where the maximum amount of information is to be extracted from the respondents. If the stimulus set is relatively large, sorting and paired comparisons are better suited for collecting similarity data. Which of these two to use will depend on characteristics of the application, such as number of stimuli and whether or not individual-level perceptual maps are desired.

Marketing Applications of MDS



(b) Multiple Horizontal Line Chart: Switching Percent x Percent Savings

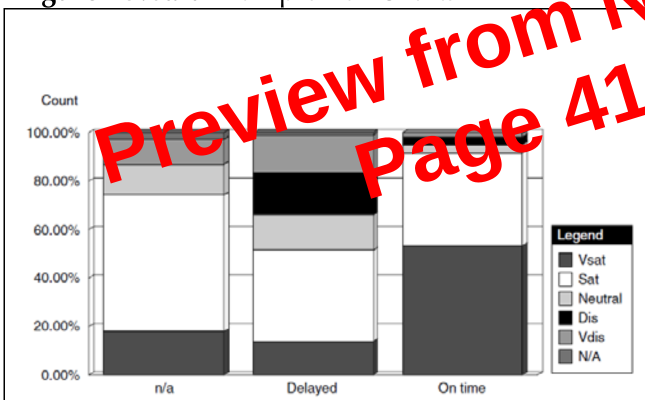


(d) Multiple Horizontal Line Chart with Horizontal Bars

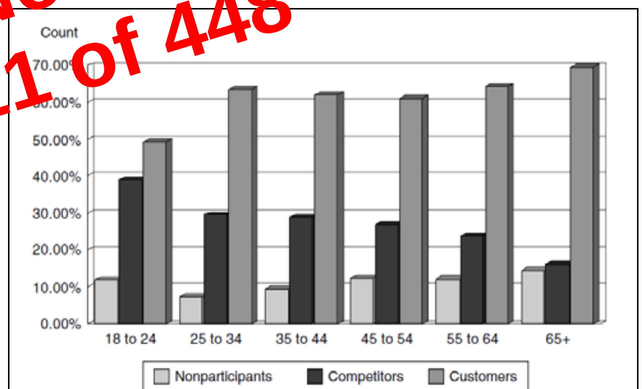
Bar Chart

A *bar chart* is another useful type of graph. Such a chart shows the relationship between a dependent variable and some independent variable at discrete levels. The stacked vertical bar chart shown in Figure 15.6a is used to show the relative satisfaction given on-time performance (*x*-axis). Figure 15.6b shows percentages associated with frequency counts of three different market segments by age categories.

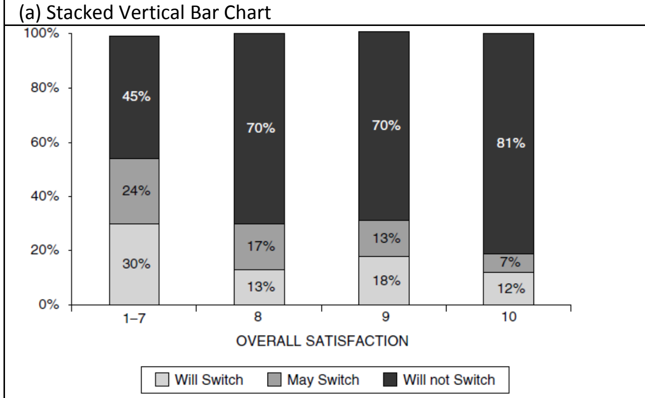
Figure 15.6a-d Example Bar Charts



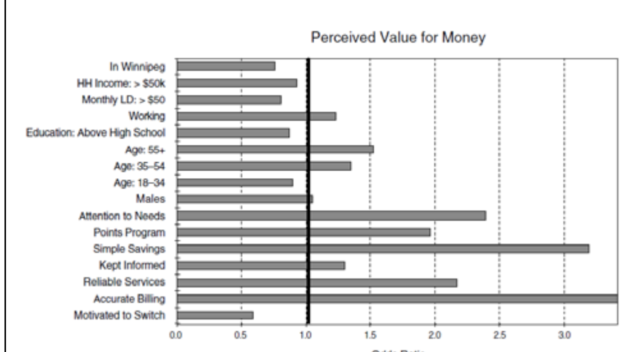
(a) Stacked Vertical Bar Chart



(b) Vertical Bar Chart: Competitive Market Share by Customer Age



(c) Stacked Vertical Bar Chart: Loyalty x Satisfaction

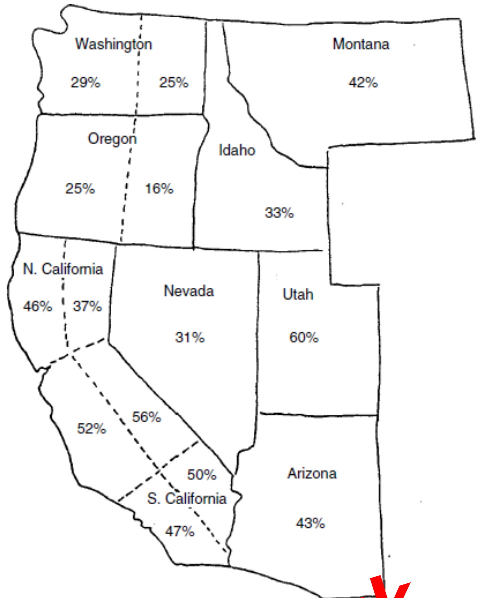


(d) Multivariable Horizontal Bar Chart with Threshold Line

Figure 15.10a-e Use of Maps and Pictograms in Displaying Research Results

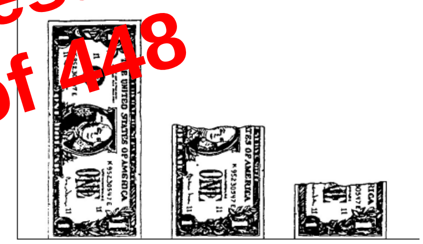
Selling Area	Percentage of All PIP Respondents Selling in the Area	Percentage of PIP Sellers With More Than 5% Sales in the Area
Area 1	78%	43%
Area 2	71%	71%
Area 3	57%	57%
Area 4	14%	0%
Area 5	57%	57%
Area 6	43%	29%

(a) Use of Maps in Plastic Irrigation Pipe Study: Selling Areas of PIP Manufacturers



(b) Use of Maps in Plastic Irrigation Pipe Study: PIP Respondent Selling Areas as a Percentage of All Respondents Reporting Sales in Area

(c) Pictogram Illustration of Number of Customers by Sales Area

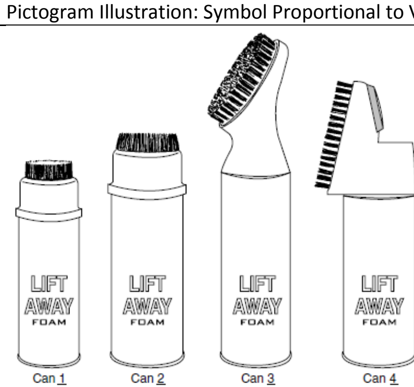


(d) Pictogram Illustration: Symbol Proportional to Value

Tc

	Most Prefer	8	6	124	82
	2nd Most Prefer	12	16	75	116
Preference	3rd Most Prefer	47	152	9	12
Without Price	Least Prefer	152	46	12	10

(e) Frequencies of Preference for New Package Designs



Pictures or drawings themselves may be useful to show what something looks like. A manufacturer of carpet cleaning agents developed a new type of foam cleaner called *Lift Away*, made especially for spot cleaning. A study was done to test four alternative aerosol packages, which used a sample of 220 housewives. Figure 15.10e shows sketches of the four package designs and frequencies of choice. It would have been difficult to describe the packages in words only. The sketches greatly aid understanding of the differences between the test items.

Table A.2 Upper Percentiles of the *t* Distribution

1– α df	.75	.90	.95	.975	.99	.995	.9995
1	1.000	3.078	6.314	12.706	31.821	63.657	636.619
2	.816	1.886	2.920	4.303	6.965	9.925	31.598
3	.765	1.638	2.353	3.182	4.541	5.841	12.941
4	.741	1.533	2.132	2.776	3.747	4.604	8.610
5	.727	1.476	2.015	2.571	3.365	4.032	6.859
6	.718	1.440	1.943	2.447	3.143	3.707	5.959
7	.711	1.415	1.895	2.365	2.998	3.499	5.405
8	.706	1.397	1.860	2.306	2.896	3.355	5.041
9	.703	1.383	1.833	2.262	2.821	3.250	4.781
10	.700	1.372	1.812	2.228	2.764	3.169	4.587
11	.697	1.363	1.796	2.201	2.718	3.106	4.437
12	.695	1.356	1.782	2.179	2.681	3.055	4.318
13	.694	1.350	1.771	2.160	2.650	3.012	4.221
14	.692	1.345	1.761	2.145	2.624	2.977	4.140
15	.691	1.341	1.753	2.131	2.602	2.947	4.073
16	.690	1.337	1.746	2.120	2.583	2.921	4.015
17	.689	1.333	1.740	2.110	2.567	2.898	3.965
18	.688	1.330	1.734	2.101	2.552	2.878	3.922
19	.688	1.328	1.729	2.093	2.339	2.861	3.883
20	.687	1.325	1.725	2.086	2.528	2.845	3.850
21	.686	1.323	1.721	2.080	2.518	2.831	3.819
22	.686	1.321	1.717	2.074	2.508	2.819	3.792
23	.685	1.319	1.714	2.069	2.500	2.807	3.767
24	.685	1.318	1.711	2.064	2.492	2.797	3.745
25	.684	1.316	1.708	2.060	2.485	2.787	3.725
26	.684	1.315	1.706	2.056	2.479	2.779	3.707
27	.684	1.314	1.703	2.052	2.473	2.771	3.690
28	.683	1.313	1.701	2.048	2.467	2.763	3.674
29	.683	1.311	1.699	2.045	2.462	2.756	3.659
30	.683	1.310	1.697	2.042	2.457	2.750	3.646
40	.681	1.303	1.684	2.021	2.423	2.704	3.551
60	.679	1.296	1.671	2.000	2.390	2.660	3.460
120	.677	1.289	1.658	1.980	2.358	2.617	3.373
∞	.674	1.282	1.645	1.960	2.326	2.576	3.291

df = degrees of freedom
SOURCE: From Fisher, R. A., & Yates, F., Statistical Tables for Biological, Agricultural, and Medical Research, 6/e, copyright © 1963. Reprinted with permission from Pearson Education, Ltd.

Table A.5 Short Table of Random Digits

9221	3452	9754	8813	6679	3081	3945	9982	1510	827
8127	9422	6665	2154	5450	7042	1481	8014	0978	865
4927	7883	9205	7590	3627	1888	1641	8781	4144	508
6213	5446	3717	0552	9623	0088	9843	0634	6831	832
5334	7757	4781	5395	3258	0296	7318	2323	2485	471
3716	5373	2907	9488	1166	1247	4451	6553	0183	394
5145	3765	9427	6725	4461	1222	8457	8517	5350	868
9851	0178	9628	7183	4597	7034	6444	4269	3567	546
3074	1224	2657	6255	2151	2313	6834	6522	7554	639
1588	1272	6511	3349	2966	8131	6845	8587	1967	646
6687	3025	2377	8365	8044	1782	0422	6664	3636	834
0637	9907	1325	1033	6821	3481	5749	6173	1494	546
7231	7647	3772	5057	0121	6369	2426	2618	2569	811
7484	4844	5451	1561	4101	5482	5315	5877	6473	786
7642	6822	8268	5671	6734	5773	6103	5138	2626	053
3371	6578	6175	6397	8748	0442	4084	6057	6865	695
2147	2406	3484	4852	3824	3382	4844	6785	9836	173
8599	7373	4168	6777	5902	3235	0784	0686	2108	841
0165	0783	5414	6605	8781	4367	4422	9073	1186	852
6838	7236	7321	7445	2493	4579	8267	1116	5165	726
6239	7121	3293	3523	8586	1748	6837	2329	6150	101
8630	9465	1743	7576	8735	7719	8041	4547	4376	674
8136	9768	0114	3089	7523	1724	3166	6376	[illegible]	[illegible]
6457	8345	2083	1333	6147	8113	5458	[illegible]	[illegible]	360
4911	0984	1290	0011	7488	7126	[illegible]	[illegible]	7131	782
8326	4220	5646	2386	6453	[illegible]	[illegible]	5781	[illegible]	346
7831	5687	6449	3576	9733	[illegible]	7275	[illegible]	[illegible]	677
9599	9382	3147	0763	[illegible]	7017	0611	[illegible]	[illegible]	261
1774	1331	7482	[illegible]	[illegible]	4427	[illegible]	2794	2153	624
0715	6621	[illegible]	[illegible]	8465	[illegible]	[illegible]	3224	0256	993
3208	[illegible]	[illegible]	0753	1886	[illegible]	8593	6527	8207	762
[illegible]	[illegible]	4449	4956	[illegible]	[illegible]	1642	4717	6252	366
[illegible]	[illegible]	2765	[illegible]	[illegible]	3083	9756	2406	5770	146
[illegible]	1877	1339	[illegible]	1243	6217	6452	2915	9888	582
6514	1122	3156	6239	3702	7011	7979	4711	5875	648
8394	5148	6209	8063	2389	1893	9125	6024	6337	845
5531	8886	8814	2202	2345	2212	3722	3606	5267	112
6271	8587	2617	8619	5073	3825	8246	5528	2517	770
2846	0853	9442	5267	4900	6451	1633	1359	4895	233
1996	0414	6748	2185	4977	3271	9411	4189	4257	087
8093	6724	3822	4761	9475	1245	3694	3453	1188	637
8688	7831	9877	4522	1423	1922	3852	1543	8419	081
1756	7154	3675	7702	0852	3859	5641	0651	9345	058
8112	8434	8807	0362	7829	1944	4644	5104	5348	422
1421	2538	3284	9826	9634	7782	8541	2381	6327	316
7298	6803	2634	8535	0604	7546	1468	5351	7794	206
2782	3515	8334	5127	4282	7832	2779	5837	3176	467
3209	9124	7177	1739	6871	7124	8845	8928	8434	900
9572	2166	0252	0571	2982	2319	3913	8976	2482	579
1877	3400	4822	3143	6211	3138	0253	4542	7575	530
0635	8728	9471	6782	8265	1065	7239	6273	6455	201
2282	4461	1485	2323	1556	2248	8216	7776	3554	655
3638	1484	5648	2673	5642	6048	9777	6035	6918	235
6938	4589	4821	5443	5515	3476	5885	2462	6393	943
7337	3261	6451	8698	0813	0249	5328	4866	1410	617
8782	5228	8387	7138	8954	5868	5456	5640	6977	857
5840	0548	4630	1764	5476	9654	7243	1354	9241	681
3307	5215	7673	3446	2861	1863	6363	4484	4684	676
1001	6776	0247	5929	9125	7368	3679	8355	3816	267
4877	4617	2257	0256	0135	4269	2567	7410	7322	877
5032	9802	9831	6230	3636	8888	2354	7093	7216	768
5230	3531	0676	9427	0702	7546	5664	9757	0353	110
3342	5442	6198	6621	7387	9691	1833	1387	1754	511

Stub-and-banner table A table that presents one dependent variable cross-tabulated by multiple independent variables.

Substantive significance An association that is statistically significant and of sufficient strength.

Sufficiency of information Degree of completeness and/or detail of information to allow a decision to be made.

Summated scale A rating scale constructed by adding scores from responses to a set of Likert scales with the purpose of placing respondents along an attitude continuum of interest. See also *Likert scale* and *multi-item scale*.

Surrogate information error Noncorrespondence of the information being sought and that required to solve the problem.

Survey A research method in which the information sought is obtained by asking questions of respondents.

Survey tracking and address books Online survey technology that uses imbedded codes to facilitate the identification and tracking of survey respondents and non-respondents.

Syndicated services Information collected and tabulated on a continuing basis by research organizations for purposes of sale to firms; data are made available to all who wish to subscribe. See *commercial data*.

Systematic error See *nonsampling error*.

Systematic sampling A probability sample where the population elements are ordered in some way and then after the first element is selected all others are chosen using a fixed interval.

Tabulation The process of sorting data into previously established categories, making initial counts of responses and using summarizing measures.

Technical report A research report that emphasizes the methods used and underlying assumptions, and presents the findings in a detailed manner.

Telephone interview Interviews that are conducted by telephone.

Telescoping A response error that occurs when a respondent reports an event happening at a time when it did not happen. It may be *forward* (report it happening more recently than it did) or *backward* (reporting it happening earlier than it did).

Testing effect The effect of a first measurement on the scores of a second measurement.

Test of independence A test of the significance of observed association involving twop or more variables.

Test-retest reliability The stability of response over time.

Thematic Apperception Test (TAT) A test consisting of one or more pictures or cartoons that depict an ambiguous situation relating to the subject being studied, and research subjects are asked to make up a story about what is happening, or the subject is asked to assume the role of a person in the situation and then describe what is happening and who the others in the scene are.