- Human relations.
- Technology.
- Training.
- Strategic planning.
- Other action areas.
- No action area identified.

How can we categorize a response suggesting a combined training-technology process? Exhibit 16-5 illustrates a combination of alternatives. By taking the categories of the first list of the action areas, it is possible to get an accurate frequency count of the joint classification possibilities for this question.

Using available software, the researcher can spend much less time coding open-ended responses and capturing categories. Software also eliminates the high cost of sending responses to outside coding firms. What used to take a coding staff several days may now be done in a few hours.

Content analysis software applies statistical algorithms to open-ended question responses. This permits stemming, aliasing, and exclusion processes. *Stemming* uses derivations of common root words to create aliases (e.g., using *searching, searches, searched,* for *search*). *Aliasing* searches for synonyms (*wise* or *smart* for *intelligent*). *Exclusion* filters out trivial words (*be, is, the, of*) in the search for meaning.[4]

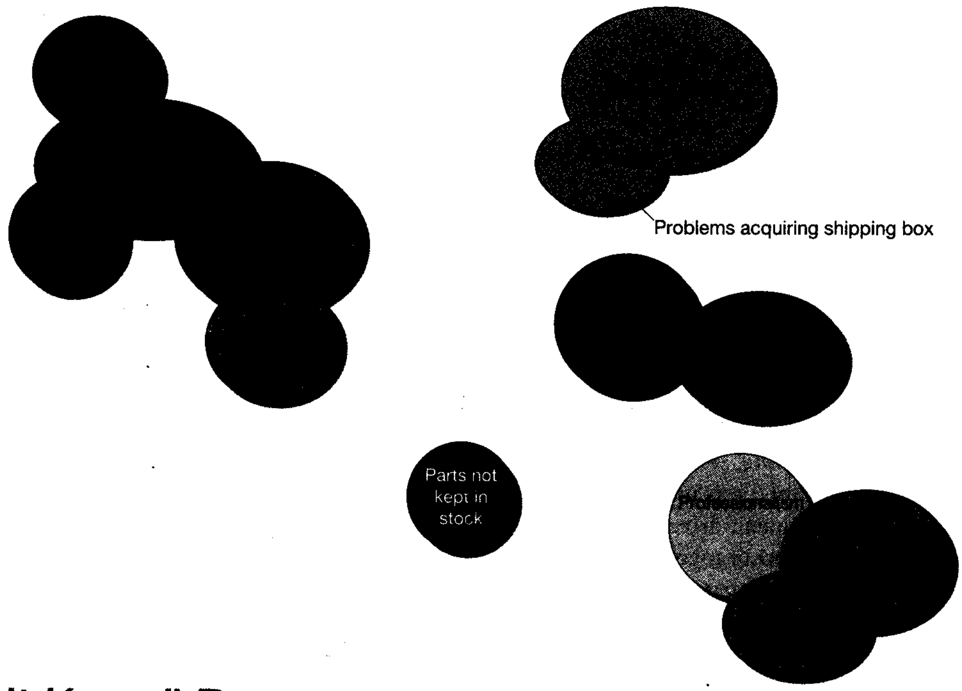
When you are using menu-driven programs, an autocategorization option creates manageable categories by clustering terms that occur together throughout the textual data set. Then, with a few keystrokes, you can modify categorization parameters and refine your results. Once your categories are consistent with the research and investigative questions, you select what you want to export to a data file or in tab-delimited format. The output, in the form of tables and plots, serves as modules for your final report. Exhibit 16-6 shows a plot produced by a content analysis of the MindWriter complaints data. The distances between pairs of terms reveal how likely it is that the terms occur together, and the colors represent categories.

> Exhibit 16-5 Open Question Coding (after revision)

Question: "How can company–customer relations be improved?"

Locus of Responsibility	Frequency (n = 100)
A. Management	
1. Sales manager	10
2. Sales process	20
3. Other	7
5. No action area identified	3
B. Salesperson	
1. Training	15
C. Customer	
1. Buying processes	12
2. Other	8
3. No action area identified	5
D. Environmental conditions	
E. Technology	20
F. Other	

> Exhibit 16-6 Proximity Plot of MindWriter Customer Complaints



"Don't Know" Responses

The **"don't know" (DK) response** presents special problems for data preparation. When the DK response group is small, it is not troublesome. But there are times when it is of major concern, and it may even be the most frequent response received. Does this mean the question that elicited this response is useless? The answer is, "It all depends." Most DK answers fall into two categories.[5] First, there is the legitimate DK response when the respondent does not know the answer. This response meets our research objectives; we expect DK responses and consider them to be useful.

In the second situation, a DK reply illustrates the researcher's failure to get the appropriate information. Consider the following illustrative questions:

1. Who developed the Managerial Grid concept?
2. Do you believe the new president's fiscal policy is sound?
3. Do you like your present job?
4. Which of the various brands of chewing gum do you believe has the best quality?
5. How often each year do you go to the movies?

It is reasonable to expect that some legitimate DK responses will be made to each of these questions. In the first question, the respondents are asked for a level of information that they often will not have. There seems to be little reason to withhold a correct answer if known. Thus, most DK answers to this question should be considered as legitimate. A DK response to the second question presents a different problem. It is not immediately clear whether the respondent is ignorant of the president's fiscal policy or knows the policy but has not made a judgment about it. The researchers should have asked two questions: In the first, they would have determined the respondent's level of awareness of fiscal policy. If the interviewee passed the awareness test, then a second question would have secured judgment on fiscal policy.

> Exhibit 16-7 Handling "Don't Know" Responses

Question: Do you have a productive relationship with your present salesperson?

Years of Purchasing	Yes	No	Don't Know
Less than 1 year	10%	40%	38%
1–3 years	30	30	32
4 years or more	60	30	30
Total	100%	100%	100%
	n = 650	n = 150	n = 200

In the remaining three questions, DK responses are more likely to be a failure of the questioning process, although some will surely be legitimate. The respondent may be reluctant to give the information. A DK response to question 3 may be a way of saying, "I do not want to answer that question." Question 4 might also elicit a DK response in which the reply translates to "This is too unimportant to talk about." In question 5, the respondents are being asked to do some calculation about a topic to which they may attach little importance. Now the DK may mean "I do not want to do that work for something of so little consequence."

Dealing with Undesired DK Responses

The best way to deal with undesired DK answers is to design better questions at the beginning. Researchers should identify the questions for which a DK response is unsatisfactory and design around it. Interviewers, however, often inherit this problem and must deal with it in the field. Several actions are then possible. First, good interviewer-respondent rapport will motivate respondents to provide more usable answers. When interviewers recognize an evasive DK response, they can repeat the question or probe for a more definite answer. The interviewer may also record verbatim any elaboration by the respondent and pass the problem on to the editor.

If the editor finds many undesired responses, little can be done unless the verbatim comments can be interpreted. Understanding the real meaning relies on clues from the respondent's answers to other questions. One way to do this is to estimate the allocation of DK answers from other data in the questionnaire. The pattern of responses may parallel income, education, or experience levels. Suppose a question concerning whether customers like their present salesperson elicits the answers in Exhibit 16-7. The correlation between years of purchasing and the "don't know" answers and the "no" answers suggests that most of the "don't knows" are disguised "no" answers.

There are several ways to handle "don't know" responses in the tabulations. If there are only a few, it does not make much difference how they are handled, but they will probably be kept as a separate category. If the DK response is legitimate, it should remain as a separate reply category. When we are not sure how to treat it, we should keep it as a separate reporting category and let the research sponsor make the decision.

Missing Data

Missing data are information from a participant or case that is not available for one or more variables of interest. In survey studies, missing data typically occur when participants accidentally skip, refuse to answer, or do not know the answer to an item on the questionnaire. In longitudinal studies, missing data may result from participants dropping out of the study, or being absent for one or more data collection periods. Missing data also occur due to researcher error, corrupted data files, and changes in the research or instrument design after data were collected from some participants, such as when variables are dropped or added. The strategy for handling

> **Exhibit 16-8** MindWriter Data Set: Missing and Out-of-Range Data

Case	1A	1B	2A	2B	2C
1	5.0	5.0	5.0	5.0	9.0
2	7.0	3.0		4.0	9.0
3	5.0	5.0	5.0	5.0	5.0
4	5.0	5.0	4.0		
5	1.0			2.0	
6	5.0	5.0	5.0	5.0	9.0
7	5.0	5.0	5.0	5.0	5.0
8	4.0	3.0	3.0	3.0	3.0
9	4.0	4.0	5.0	5.0	5.0
10	4.0	5.0	.	4.0	5.0
11	2.0	5.0	4.0	4.0	5.0
12	6.0	4.0	3.0	3.0	4.0
13	5.0	5.0		3.0	5.0
14	5.0	5.0	5.0	5.0	5.0
15	5.0	4.0	5.0	5.0	4.0
Valid	15	14	11	14	13
Missing	0	1	4	1	2
Mean	4.53	4.50	4.45	4.14	5.61
Range	6	2	2	3	6
Minimum	1	3	3	2	3
Maximum	7	5	5	5	9

missing data consists of a two-step process: the researcher first explores the pattern of missing data to determine the mechanism for *missingness* (the probability that a value is missing rather than observed) and then selects a missing-data technique.

Examine the sample distribution of variables from the MindWriter dataset shown in Exhibit 16-8. These data were collected on a five-point interval scale. There are no missing data in variable 1A, although it is apparent that a range of 6 and a maximum value of 7 invalidate the calculated mean or average score. Variables 1B and 2B have one case missing but values that are within range. Variable 2A is missing four cases, or 27 percent of its data points. The last variable, 2C, has a range of 6, two missing values, and three values coded as "9." A "9" is often used as a DK or missing-value code when the scale has a range of less than 9 points. In this case both blanks and 9s are present—a coding concern. Notice that the fifth respondent answered only two of the five questions and the second respondent had two miscoded answers and one missing value. Finally, using descriptive indexes of shape, discussed in Appendix 16a, you can find three variables that depart from the symmetry of the normal distribution. They are skewed (or pulled) to the left by a disproportionately small number of 1s and 2s. And one variable's distribution is peaked beyond normal dimensions. We have just used the minimum and maximum values, the range, and the mean and have already discovered errors in coding, problems with respondent answer patterns, and missing cases.

Mechanisms for Missing Data

In order to select a missing-data technique, the researcher must first determine what caused the data to be missing. There are three basic mechanisms for this: data missing completely at random (MCAR); data missing at random (MAR); and data not missing at random (NMAR). If the probability of missingness for a particular variable is dependent on neither the variable itself nor any other variable in the data set, then data are MCAR. Data are considered MAR if the probability of missingness for a particular variable is dependent on another variable but not itself when other variables are held constant. The practical significance of this distinction is that the proper missing-data technique can be selected that will minimize bias in subsequent analyses. The third type of mechanism, NMAR, occurs when data are not missing completely at random and they are not predictable from other variables in the data set. Data NMAR are considered *nonignorable* and must be treated on an improvised basis.

Missing-Data Techniques

Three basic types of techniques can be used to salvage data sets with missing data: (1) listwise deletion, (2) pairwise deletion, and (3) replacement of missing values with estimated scores. *Listwise deletion,* or complete case analysis, is perhaps the simplest approach, and is the default option in most statistical packages like SPSS and SAS. With this method, cases are deleted from the sample if they have missing values on any of the variables in the analysis. Listwise deletion is most appropriate when data are MCAR. In this situation, no bias will be introduced because the subsample of complete cases is essentially a random sample of the original sample. However, if data are MAR but not MCAR, then a bias may be introduced, especially if a large number of cases are deleted. For example, if men were more likely than women to be responsible for missing data on the variable *shopping preference,* then the results would be biased toward women's shopping preferences.

Pairwise deletion, also called available case analysis, assumes that data are MCAR. In the past, this technique was used frequently with linear models that are functions of means, variances, and covariances. Missing values would be estimated using all cases that had data for each variable or pair of variables in the analysis. Today most experts caution against pairwise deletion, and recommend alternative approaches.

The replacement of missing values with estimated values includes a variety of techniques. This option generally assumes that data are MAR, since the missing values on one variable are predicted from observed values on another variable. A common option available on many software packages is the replacement of missing values with a mean or other central tendency score. This is a simple approach, but has the disadvantage of reducing the variability in the original data, which can cause bias. Another option is to use a regression or likelihood-based method. Such techniques are found in specialty software packages and the procedures for using them are beyond the scope of this text.

> Data Entry

Data entry converts information gathered by secondary or primary methods to a medium for viewing and manipulation. Keyboarding remains a mainstay for researchers who need to create a data file immediately and store it in a minimal space on a variety of media. However, researchers have profited from more efficient ways of speeding up the research process, especially from bar coding and optical character and mark recognition.

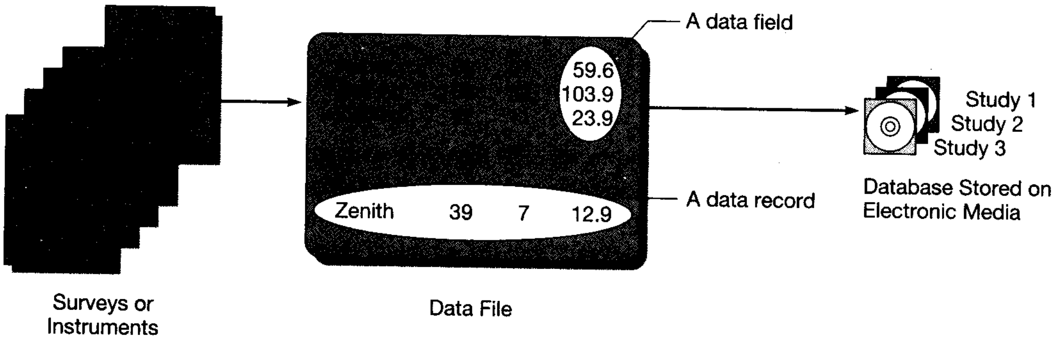
Alternative Data Entry Formats

Keyboarding

A full-screen editor, where an entire data file can be edited or browsed, is a viable means of data entry for statistical packages like SPSS or SAS. SPSS offers several data entry products, including Data Entry Builder™,

> Exhibit 16-9 Data Fields, Records, Files, and Databases

Data fields represent single elements of information (e.g., an answer to a particular question) from all participants in a study. Data fields can contain numeric, alphabetic, or symbolic information. A **record** is a set of data fields that are related to one case or participant (e.g., the responses to one completed survey). Records represent rows in a data file or spreadsheet program worksheet. **Data files** are sets of records (e.g., responses from all participants in a single study) that are grouped together for storage on diskettes, disks, tapes, CD-ROM, or optical disks. *Databases* are made up of one or more data files that are interrelated. A database might contain all customer survey information collected quarterly for the last 10 years.



which enables the development of forms and surveys, and Data Entry Station™, which gives centralized entry staff, such as telephone interviewers or online participants, access to the survey. Both SAS and SPSS offer software that effortlessly accesses data from databases, spreadsheets, data warehouses, or data marts.

Database Development For large projects, database programs serve as valuable data entry devices. A **database** is a collection of data organized for computerized retrieval. Programs allow users to define data fields and link files so that storage, retrieval, and updating are simplified. The relationship between *data fields, records, files,* and *databases* is illustrated in Exhibit 16-9. A company's orders serve as an example of a database. Ordering information may be kept in several files: salesperson's customer files, customer financial records, order production records, and order shipping documentation. The data are separated so that authorized people can see only those parts pertinent to their needs. However, the files may be linked so that when, say, a customer changes his or her shipping address, the change is entered once and all the files are updated. Another database entry option is e-mail data capture. It has become popular with those using e-mail-delivered surveys. The e-mail survey can be delivered to a specific respondent whose e-mail address is known. Questions are completed on the screen, returned via e-mail, and incorporated into a database.[6] An intranet can also capture data. When participants linked by a network take an online survey by completing a database form, the data are captured in a database in a network server for later or real-time analysis.[7] ID and password requirements can keep unwanted participants from skewing the results of an online survey.

Researchers consider database entry when they have large amounts of potentially linked data that will be retrieved and tabulated in different ways over time. Another application of a database program is as a "front-end" entry mechanism. A telephone interviewer may ask the question "How many children live in your household?" The computer's software has been programmed to accept any answer between 0 and 20. If a "P" is accidentally struck, the program will not accept the answer and will return the interviewer to the question. With a precoded online instrument, some of the editing previously discussed is done by the program. In addition, the program can be set for automatic conditional branching. In the example, an answer of 1 or greater causes the program to prompt the questioner to ask the ages of the children. A 0 causes the age question to be automatically skipped. Although this option is available whenever interactive computing is used, front-end processing is typically done within the database design. The database will then store the data in a set of linked files that allow the data to be easily sorted. Descriptive statistics and tables—the first steps in exploring data—are readily generated from within the database.

> Exhibit 16-10 Data Entry Using Spreadsheets

Each row is a record (a single participant's responses). Each column is a variable measured in the survey. In this survey, questions 1, 3, and 5 are nominal variables that have two response categories. Question 6 uses multiple columns as it is a multipart rating question using a 1-to-5 scale. This is a typical way of coding variables in a spreadsheet before they are imported by SPSS (assuming you are using a spreadsheet instead of the SPSS Data Editor to start your study). Note that each participant is assigned an identification number (case ID). After running preliminary frequencies, having a case ID data field enables you to quickly find and correct suspect data like odd value codes or missing cases.

Case ID	Q1	Q2	Q3	Q4	Q5	Q6a	Q6b	Q6c	Q6d	Q6e	Q6f
0001	1	2	1	10	2	1	2	1	1	4	4
0002	2	5	2	7	1	2	2	3	2	4	5
0003	1	2	1	6	2	2	4	3	4	4	4
0004	1	2	1	1	1	3	4	4	4	5	4
0005	2	6	2	8	2	3	5	4	2	5	1
0006	2	1	2	8	2	3	5	2	2	3	1
0007	1	3	1	8	1	2	5	3	5	3	3
0008	2	4	2	5	2	3	3	4	5	1	3
0009	1	2	1	9	1	3	2	4	5	2	5
0010	2	2	2	9	2	4	2	5	5	3	5
0011	2	5	2	9	1	4	1	1	3	1	5
0012	1	2	1	9	1	2	2	2	3	2	2
0013	2	1	2	3	2	5	3	3	4	2	1
0014	1	6	1	2	2	3	4	4	5	5	2
0015	2	4	2	3	1	1	4	3	1	5	3
0016	2	3	2	4	2	5	5	5	2	5	4
0017	1	3	1	6	1	5	5	2	1	1	4
0018	2	3	2	5	2	5	5	2	2	2	3

Spreadsheet Spreadsheets are a specialized type of database for data that need organizing, tabulating, and simple statistics. They also offer some database management, graphics, and presentation capabilities. Data entry on a **spreadsheet** uses numbered rows and lettered columns with a matrix of thousands of cells into which an entry may be placed. Spreadsheets allow you to type numbers, formulas, and text into appropriate cells. Many statistics programs for personal computers and also charting and graphics applications have data editors similar to the Excel spreadsheet format shown in Exhibit 16-10. This is a convenient and flexible means for entering and viewing data.

Optical Recognition

If you use a PC image scanner, you probably are familiar with **optical character recognition (OCR)** programs that transfer printed text into computer files in order to edit and use it without retyping. There are other, related applications. **Optical scanning** of instruments—the choice of testing services—is efficient for researchers. Examinees darken small circles, ellipses, or spaces between sets of parallel lines to indicate their answers. A more flexible format, **optical mark recognition (OMR)** uses a spreadsheet-style interface to read and process user-created forms. Optical scanners process the marked-sensed questionnaires and store the answers in a file. This method, most often associated with standardized and preprinted forms, has been adopted by researchers for data entry and preprocessing due to its speed (10 times faster than keyboarding), cost savings on data entry, convenience in charting and reporting data, and improved accuracy. It reduces the number of times data are handled, thereby reducing the number of errors that are introduced.

Other techniques include direct-response entry, of which voting procedures used in several states are an example. With a specially prepared punch card, citizens cast their votes by pressing a pen-shaped instrument

In 2004 Princess Cruises had 15 ships and 30,000 berths sailing 7 to 72 days to more than 6 continents on more than 150 itineraries. Princess carries more than 700,000 passengers each year and processes 245,000 customer satisfaction surveys each year—distributed to each cabin on the last day of each cruise. Princess uses scannable surveys rather than human data entry for one reason: in the 1-week to 10-day analysis lag created by human data entry, 10 cruises could be completed with another 10 under way. For a business that prides itself on customer service, not knowing about a problem could be enormously damaging. Princess has found that scannable surveys generate more accurate data entry, while reducing processing and decision-response time—critical time in the cruise industry. **www.princesscruises.com**

against the card next to the preferred candidate. This opens a small hole in a specific column and row of the card. The cards are collected and placed directly into a card reader. This method also removes the coding and entry steps. Another governmental application is the 1040EZ form used by the Internal Revenue Service. It is designed for computerized number and character recognition. Similar character recognition techniques are employed for many forms of data collection. Again, both approaches move the response from the question to data analysis with little handling.

Voice Recognition

The increase in computerized random dialing has encouraged other data collection innovations. **Voice recognition** and voice response systems are providing some interesting alternatives for the telephone interviewer. Upon getting a voice response to a randomly dialed number, the computer branches into a questionnaire routine. These systems are advancing quickly and will soon translate recorded voice responses into data files.

Digital

Telephone keypad response, frequently used by restaurants and entertainment venues to evaluate customer service, is another capability made possible by computers linked to telephone lines. Using the telephone keypad (touch-tone), an invited participant answers questions by pressing the appropriate number. The computer captures the data by decoding the tone's electrical signal and storing the numeric or alphabetic answer in a data file. While not originally designed for collecting survey data, each of the software components within Microsoft Office XP includes advanced speech recognition functionality, enabling people to enter and edit data by speaking into a microphone.[8]

>snapshot

VNS: This Hit Was the Knockout

After the fiasco of the 2000 presidential election when Voter News Service (VNS) exit polls resulted in miscalled election results, withdrawal of concession speeches, and a month of Florida recounts, VNS invested more than $10 million to upgrade its computing technology, including the installation of a voice recognition data entry system. But during the November 2002 congressional elections, the VNS consortium owners—ABC, CBS, CNN, Fox News, NBC, and the Associated Press—had to abandon their use of exit polls midstream. The technology overhaul had attempted (unsuccessfully) to merge two operating systems, get those systems to successfully integrate multiple databases, and launch the voice recognition data entry system without allowing sufficient testing time. Many of the more than 30,000 interviewers collecting exit-poll information were disconnected before they could finish inputting data over the phone. Others couldn't gain entry to the new system at all. Early exit polls proved just as inaccurate as the polls two years earlier. For example, in one highly visible contest, exit polls showed Elizabeth Dole being defeated for the North Carolina senatorial race when she was actually winning. Dole ultimately claimed her seat with a decisive 200,000-vote margin. This second right hook proved a knockout for VNS; on January 13, 2003, the news-research consortium conceded defeat and disbanded VNS.

Field interviewers can use mobile computers or notebooks instead of clipboards and pencils. With a built-in communications modem, wireless LAN, or cellular link, their files can be sent directly to another computer in the field or to a remote site. This lets supervisors inspect data immediately or simplifies processing at a central facility. This is the technology that Nielsen Media is using with its portable People Meter.

Bar Code Since adoption of the Universal Product Code (UPC) in 1973, the bar code has developed from a technological curiosity to a business mainstay. After a study by McKinsey & Company, the Kroger grocery chain pilot-tested a production system and bar codes became ubiquitous in that industry.[9]

Bar-code technology is used to simplify the interviewer's role as a data recorder. When an interviewer passes a bar-code wand over the appropriate codes, the data are recorded in a small, lightweight unit for translation later. In the large-scale processing project Census 2000, the Census Data Capture Center used bar codes to identify residents. Researchers studying magazine readership can scan bar codes to denote a magazine cover that is recognized by an interview participant.

The **bar code** is used in numerous applications: point-of-sale terminals, hospital patient ID bracelets, inventory control, product and brand tracking, promotional technique evaluation, shipment tracking, marathon runners, rental car locations (to speed the return of cars and generate invoices), and tracking of insects' mating habits. The military uses 2-foot-long bar codes to label boats in storage. The codes appear on business documents, truck parts, and timber in lumberyards. Federal Express shipping labels use a code called *Codabar.* Other codes, containing letters as well as numbers, have potential for researchers.

On the Horizon

Even with these time reductions between data collection and analysis, continuing innovations in multimedia technology are being developed by the personal computer business. The capability to integrate visual images, streaming video, audio, and data may soon replace video equipment as the preferred method for recording an experiment, interview, or focus group. A copy of the response data could be extracted for data analysis, but the audio and visual images would remain intact for later evaluation. Although technology will never replace researcher judgment, it can reduce data-handling errors, decrease time between data collection and analysis, and help provide more usable information.

>summary

1 The first step in data preparation is to edit the collected raw data to detect errors and omissions that would compromise quality standards. The editor is responsible for making sure the data are accurate, consistent with other data, uniformly entered, and ready for coding. In survey work, it is common to use both field and central editing.

2 Coding is the process of assigning numbers and other symbols to answers so that we can classify the responses into categories. Categories should be appropriate to the research problem, exhaustive of the data, mutually exclusive, and unidimensional. The reduction of information through coding requires that the researcher design category sets carefully, using as much of the data as possible. Codebooks are guides to reduce data entry error and serve as a compendium of variable locations and other information for the analysis stage. Software developments in survey construction and design include embedding coding rules that screen data as they are entered, identifying data that are not entered correctly.

3 Closed questions include scaled items and other items for which answers are anticipated. Precoding of closed items avoids tedious completion of coding sheets for each response. Open-ended questions are more difficult to code since answers are not prepared in advance, but they do encourage disclosure of complete information. A systematic method for analyzing open-ended questions is content analysis. It uses preselected sampling units to produce frequency counts and other insights into data patterns.

4 "Don't know" replies are evaluated in light of the question's nature and the respondent. While many DKs are legitimate, some result from questions that are ambiguous or from an interviewing situation that is not motivating. It is better to report DKs as a separate category unless there are compelling reasons to treat them otherwise. Missing data occur when respondents skip, refuse to answer, or do not know the answer to a questionnaire item, drop out of the study, or are absent for one or more data collection periods. Researcher error, corrupted data files, and changes to the instrument during administration also produce missing data. Researchers handle missing data by first exploring the data to discover the nature of the pattern and then selecting a suitable technique for replacing values by deleting cases (or variables) or estimating values.

5 Data entry is accomplished by keyboard entry from precoded instruments, optical scanning, real-time keyboarding, telephone pad data entry, bar codes, voice recognition, OCR, OMR, and data transfers from electronic notebooks and laptop computers. Database programs, spreadsheets, and editors in statistical software programs offer flexibility for entering, manipulating, and transferring data for analysis, warehousing, and mining.

>keyterms

bar code 459
codebook 444
coding 443
content analysis 448
data entry 455
data field 456
data file 456
data preparation 440
database 456
"don't know" (DK) response 452
editing 441
missing data 453
optical character recognition (OCR) 457
optical mark recognition (OMR) 457
optical scanning 457
precoding 444
record 456
spreadsheet 457
voice recognition 458

>discussionquestions

Terms in Review

1 Define or explain:
   a Coding rules.
   b Spreadsheet data entry.
   c Bar codes.
   d Precoded instruments.
   e Content analysis.
   f Missing data.

g Optical mark recognition.

2 How should the researcher handle "don't know" responses?

Making Research Decisions

3 A problem facing shoe store managers is that many shoes eventually must be sold at markdown prices. This prompts us to conduct a mail survey of shoe store managers in which we ask, "What methods have you found most successful for reducing the problem of high markdowns?" We are interested in extracting as much information as possible from these answers to better understand the full range of strategies that store managers use. Establish what you think are category sets to code 500 responses similar to the 14 given below. Try to develop an integrated set of categories that reflects your theory of markdown management. After developing the set, use it to code the 14 responses.

a Have not found the answer. As long as we buy style shoes, we will have markdowns. We use PMs on slow merchandise, but it does not eliminate markdowns. (*PM* stands for "push-money"—special item bonuses for selling a particular style of shoe.)

b Using PMs before too old. Also reducing price during season. Holding meetings with salespeople indicating which shoes to push.

c By putting PMs on any slow-selling items and promoting same. More careful check of shoes purchased.

d Keep a close watch on your stock, and mark down when you have to—that is, rather than wait, take a small markdown on a shoe that is not moving at the time.

e Using the PM method.

f Less advance buying—more dependence on in-stock shoes.

g Sales—catch bad guys before it's too late and close out.

h Buy as much good merchandise as you can at special prices to help make up some markdowns.

i Reducing opening buys and depending on fill-in service. PMs for salespeople.

j Buy more frequently, better buying, PMs on slow-moving merchandise.

k Careful buying at lowest prices. Cash on the buying line. Buying closeouts, FDs, overstock, "cancellations." (*FD* stands for "factory-discontinued" style.)

l By buying less "chanceable" shoes. Buy only what you need, watch sizes, don't go overboard on new fads.

m Buying more staple merchandise. Buying more from fewer lines. Sticking with better nationally advertised merchandise.

n No successful method with the current style situation. Manufacturers are experimenting, the retailer takes the markdowns—cuts gross profit by about 3 percent—keep your stock at lowest level without losing sales.

4 Select a small sample of class members, work associates, or friends and ask them to answer the following in a paragraph or two: "What are your career aspirations for the next five years?" Use one of the four basic units of content analysis to analyze their responses. Describe your findings as frequencies for the unit of analysis selected.

Bringing Research to Life

5 What data preparation process was Jason doing during data entry?

6 Data entry followed data collection in the research profiled during the opening vignette. What concerned Jason about this process?

From Concept to Practice

7 Choose one of the cases on your text CD that has an instrument (check the Case Abstracts section for a listing of all cases and an abstract for each). Code the instrument for data entry.

>wwwexercises

1 See what the next generation of qualitative research analysis can do. Visit the QRS Web site and take a product tour of XSight (**http://www.qsr.com.au/products/productoverview/XSight.htm**).

2 Visit the Internet home page of three of the world's biggest research companies (you'll find several of them mentioned in Chapter 1). Do a content analysis of the three home pages. Be sure to look at all formats of content—text, pictures, video, and audio—and all four types of content: syntactical, referential, propositional, and thematic. How will you categorize the data? How will you create a data record for each company? What content elements are common to all? What elements are unique to a particular research company?

>cases*

Agri Comp

Inquiring Minds Want to Know—NOW!

Mastering Teacher Leadership

NCRCC: Teeing Up and New Strategic Direction

NetConversions Influences Kelley Blue Book

* All cases appear on the text CD; you will find abstracts of these cases in the Case Abstracts section of the text.

>appendix16a

Describing Data Statistically

In the first part of the chapter, we discussed how responses from participants are edited, coded, and entered. Creating numerical summaries of this process provides valuable insights to analysts about their effectiveness. In this appendix, we review concepts from your introductory statistics course that offer descriptive tools for cleaning data, discovering problems, and summarizing distributions. A distribution (of data) is an array of value counts from lowest to highest value of a variable, resulting from the tabulation of incidence. Descriptive statistical measures are used to depict the center, spread, and shape of distributions and are helpful as preliminary tools for data description. We will define these measures and describe their use as *descriptive statistics* after introducing a sample data set and an overview of basic concepts.

Reviewing Statistical Concepts

The LCD (liquid crystal display) TV market is an interesting market to watch because of the changes in technology and marketing. Currently the major players in this market are Sharp, LG Electronics/Zenith, Samsung, Sony, Dell, and Panasonic. Only a few other brands earn a noticeable market share. Sharp products currently represent the largest percentage of unit sales. Let's assume we are interested in evaluating annual unit sales increases of several manufacturers. We survey nine manufacturers and we find a *frequency distribution* (an ordered array of all values for a variable) of annual percentage of unit sales increases: 5, 6, 6, 7, 7, 7, 8, 8, 9. From these unit sales scores, we construct a table for arraying the data. It presents value codes from lowest to highest value, with columns for count, percent, percent for missing values, and cumulative percent. An example is presented in Exhibit 16a-1.

The table arrays data by assigned numerical value, in this case the actual percentage unit sales increase recorded (far left column). To discover how many manufacturers were in each unit sales increase category, you would read the frequency column. For example, at the intersection of the frequency column and the second row there are two companies that posted a 6 percent annual unit sales increase. In the percentage column, you see what percentage of TV manufacturers in the survey gave a response for each level of unit sales increase. The three manufacturers who had unit sales increases of 7 percent represent 33.3 percent of the total number of manufacturers surveyed (3/9 × 100). The cumulative percentage reveals the number of manufacturers that provided a response and *any others that preceded it* in the table. For this example, LCD TV percentage unit sales increases between 5 and 7 percent represent 66.7 percent. The cumulative percentage column is helpful primarily when the data have an underlying order. If, in part B, we create a code for source of origin (foreign = 1, domestic = 2) to each of the nine LCD TV manufacturers, the cumulative percentage column would provide the proportion. The *proportion* is the percentage of elements in the distribution that met a criterion. In this case, the criterion is the origin of manufacture.

In Exhibit 16a-2, the bell-shaped curve that is superimposed on the distribution of annual unit sales increases (percent) for LCD TV manufacturers is called the *normal distribution*. The distribution of values for any variable that has a normal distribution is governed by a mathematical equation. This distribution is a symmetrical curve and reflects a frequency distribution of many natural phenomena such as the height of people of a certain gender and age.

Many variables of interest that researchers will measure will have distributions that approximate a *standard normal distribution*. A standard normal distribution is a special case of the normal distribution in which all values are given standard scores. This distribution has a mean of 0 and a standard deviation of 1. For example, a manufacturer that had an annual unit sales increase of 7 percent would be given a standard score of zero since 7 is the mean of the LCD TV distribution. A *standard score* (or *Z score*) tells you how many units a case (a manufacturer in this example) is above or below the mean. The Z score, being standardized, allows

> Exhibit 16a-1 Annual Percentage Unit Sales Increases for LCD TV Manufacturers

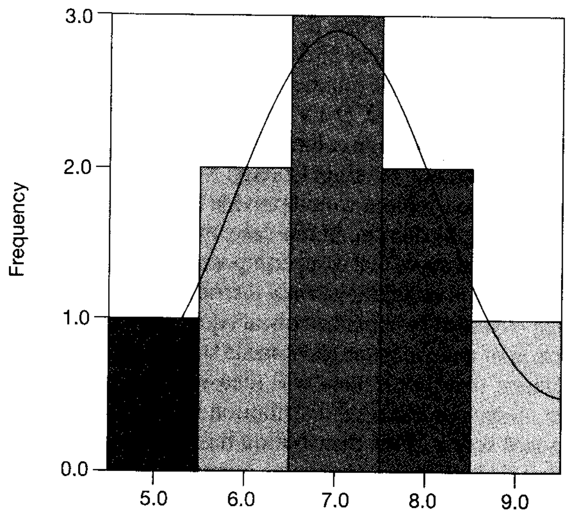
A

Unit Sales Increase (%)	Frequency	Percentage	Cumulative Percentage
5	1	11.1	11.1
6	2	22.2	33.3
7	3	33.3	66.7
8	2	22.2	88.9
9	1	11.1	100.0
Total	9	100.0	

B

Company Origin	Unit Sales Increase (%)	Frequency	Percentage	Cumulative Percentage
Origin, foreign (1)	6	1	11.1	11.1
	7	2	22.2	33.3
	8	2	22.2	55.5
Origin, domestic (2)	5	1	11.1	[illegible]
	6	1	11.1	[illegible]
	7	1	11.1	[illegible]
	9	1	11.1	[illegible]
	Total	9	100.0	

> Exhibit 16a-2 Histogram of Annual Unit Sales Increase (%)



us to compare the results of different normal distributions, something we do frequently in research. Assume that Zenith has an annual unit sales increase of 9 percent. To calculate a standard score for this manufacturer, you would find the difference between the value and the mean and divide by the standard deviation of the distribution shown in Exhibit 16a-1.

$$\text{Zenith's standard score} = \frac{\text{value} - \text{mean}}{\text{standard deviation}} = \frac{9 - 7}{1.22} = 1.64$$

The standard normal distribution, shown in part A of Exhibit 16a-3, is a standard of comparison for describing distributions of sample data. It is used with inferential statistics that assume normally distributed variables.

We will come back to this exhibit in a moment. Now let's review some descriptive tools that reveal the important characteristics of distributions. The characteristics of central tendency, variability, and shape are useful tools for summarizing distributions. Their definitions, applications, and formulas fall under the heading of *descriptive statistics.* The definitions will be familiar to most readers.

Measures of Central Tendency

Summarizing information such as that from our collected data of LCD TV manufacturers often requires the description of "typical" values. Suppose we want to know the typical percentage unit sales increase for these companies. We might define *typical* as the average response (mean); the middle value, when the distribution is sorted from lowest to highest (median); or the most frequently occurring value (mode). The common measures of *central tendency* (or center) include the mean, median, and mode.

The *mean* is calculated by the formula below:

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

For the unit sales increase variable, the distribution of responses is 5, 6, 6, 7, 7, 7, 8, 8, 9. The arithemetic average, or mean (the sum of the nine values divided by 9), is

$$\frac{5 + 6 + 6 + 7 + 7 + 8 + 8 + 8 + 9}{9} = 7 \text{ (an average 7\% unit sales increase)}$$

The *median* is the midpoint of the distribution. Half of the observations in the distribution fall above and the other half fall below the median. When the distribution has an even number of observations, the median is the average of the two middle scores. The median is the most appropriate locator of center for ordinal data and has resistance to extreme scores, thereby making it a preferred measure for interval and ratio data when their distributions are not normal. The median is sometimes symbolized by *M* or *mdn.*

From the sample distribution for the percentage unit sales increase variable, the median of the nine values is 7:

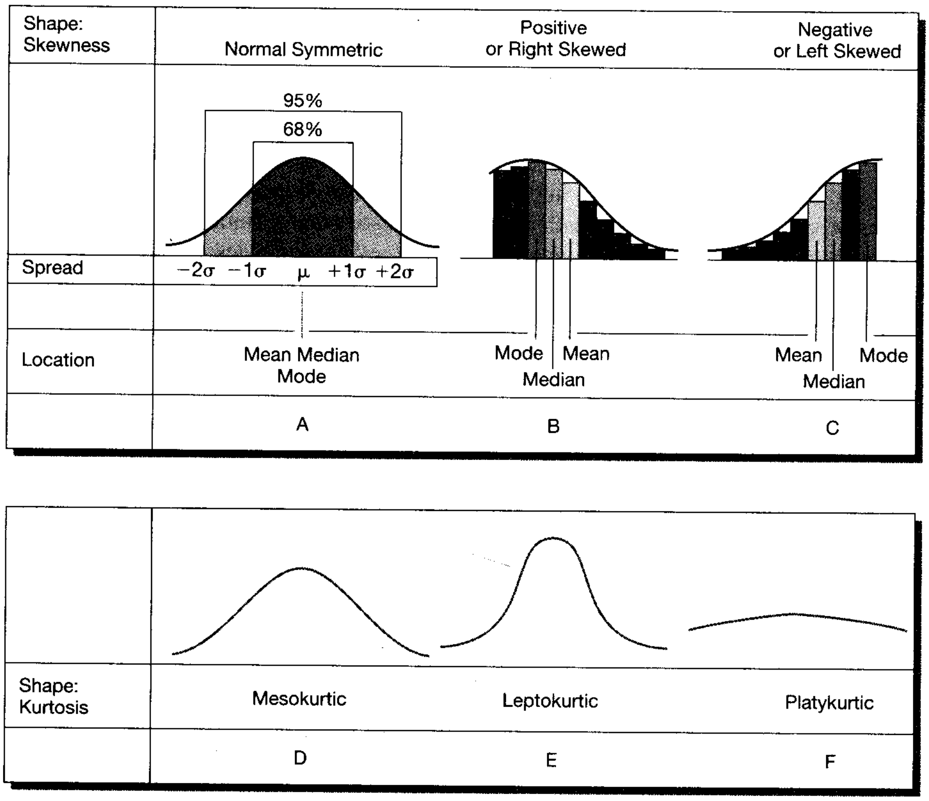
5 6 6 7 **7** 7 8 8 9

If the distribution had 10 values, the median would be the average of the values for the fifth and sixth cases.

The *mode* is the most frequently occurring value. There may be *more than one* mode in a distribution. When there is more than one score that has the highest yet equal frequency, the distribution is bimodal or multimodal. There may be *no* mode in a distribution if every score has an equal number of observations. The mode is the location measure of central tendency for nominal data and a point of reference along with the median and mean for examining spread and shape of distributions. In our LCD TV percentage unit sales increase example, the most frequently occurring value is 7. As revealed in the frequency distribution in Exhibit 16a-3, there are three companies that have unit sales increases of 7 percent.

Notice in Exhibit 16a-3, part A, that the mean, median, and mode are the same in a normal distribution. When these measures of central tendency diverge, the distribution is no longer normal.

> Exhibit 16a-3 Characteristics of Distributions



Measures of Variability

The common measures of *variability,* alternatively referred to as *dispersion* or *spread,* are the variance, standard deviation, range, interquartile range, and quartile deviation. They describe how scores cluster or scatter in a distribution.

The *variance* is a measure of score dispersion about the mean. If all the scores are identical, the variance is 0. The greater the dispersion of scores, the greater the variance. Both the variance and the standard deviation are used with interval and ratio data. The symbol for the sample variance is s^2, and that for the population variance is the Greek letter sigma, squared (σ^2). The variance is computed by summing the squared distance from the mean for all cases and dividing the sum by the total number of cases minus 1:

$$\text{Variance} = s^2 = \frac{\text{sum of the squared distances from the mean for all cases}}{(\text{number of cases} - 1)}$$

$$s^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n - 1}$$

For the percentage unit sales increase variable, we would compute the variance as

$$s^2 = \frac{(5-7)^2 + (6-7)^2 + (6-7)^2 + (7-7)^2 + (7-7)^2 + (7-7)^2 + (8-7)^2 + (8-7)^2 + (9-7)^2}{}$$

$$= 1.5$$

The *standard deviation* summarizes how far away from the average the data values typically are. It is perhaps the most frequently used measure of spread because it improves interpretability by removing the variance's square and expressing the deviations in their original units (e.g., sales in dollars, not dollars squared). It is also an important concept for descriptive statistics because it reveals the amount of variability within the data set. Like the mean, the standard deviation is affected by extreme scores. The symbol for the sample standard deviation is s, and that for a population standard deviation is σ. Alternatively, it is labeled *std. dev.* You can calculate the standard deviation by taking the square root of the variance:

$$s = \sqrt{s^2}$$

The standard deviation for the percentage unit sales increase variable in our example is 1.22:

$$1.22 = \sqrt{1.5}$$

The *range* is the difference between the largest and smallest scores in the distribution. The percentage annual unit sales increase variable has a range of 4 ($9 - 5 = 4$). Unlike the standard deviation, the range is computed from only the minimum and maximum scores; thus, it is a very rough measure of spread. With the range as a point of comparison, it is possible to get an idea of the homogeneity (small std. dev.) or heterogeneity (large std. dev.) of the distribution. For a homogeneous distribution, the ratio of the range to the standard deviation should be between 2 and 6. A number above 6 would indicate a high degree of heterogeneity. In the percentage unit sales increase example, the ratio is $4/1.22 = 3.28$. The range provides useful but limited information for all data. It is mandatory for ordinal data.

The *interquartile range (IQR)* is the difference between the first and third quartiles of the distribution. It is also called the *midspread.* Ordinal or ranked data use this measure in conjunction with the median. It is also used with interval and ratio data when asymmetrical distributions are suspected or for exploratory analysis. Recall the following relationships: The minimum value of the distribution is the 0 percentile; the maximum, the 100th percentile. The first quartile (Q_1) is the 25th percentile; the median, Q_2, is the 50th percentile. The third quartile (Q_3) is the 75th percentile. For the percentage unit sales increase data, the quartiles are

$$\begin{array}{ccccccccc} 5 & 6 & 6 & 7 & 7 & 7 & 8 & 8 & 9 \\ \hline & & Q_1 & & Q_2 & & Q_3 & & Q_4 \end{array}$$

The quartile deviation, or semi-interquartile range, is expressed as

$$Q = \frac{Q_1 - Q_3}{2}$$

The *quartile deviation* is always used with the median for ordinal data. It is helpful for interval and ratio data when the distribution is stretched (or skewed) by extreme values. In a normal distribution, the median plus one quartile deviation (Q) on either side encompasses 50 percent of the observations. Eight Qs cover approximately the range. Q's relationship with the standard deviation is constant ($Q = .6745s$) when scores are normally distributed. For our annual percentage unit sales increase example, the quartile deviation is 1 [$(6 - 8)/2 = 1$].

Measures of Shape

The measures of shape, skewness and kurtosis, describe departures from the symmetry of a distribution and its relative flatness (or peakedness), respectively. They use deviation scores ($X - \overline{X}$). *Deviation scores* show us how far any observation is from the mean. The company that posted a percentage unit sales increase of 9 has a deviation score of 2 ($9 - 7$). The measures of shape are often difficult to interpret when extreme scores are in the distribution. Generally, shape is best communicated through visual displays. (Refer to the graphics in Exhibit 16a-3, parts B through F.) From a practical standpoint, the calculation of skewness and kurtosis is easiest with spreadsheet or statistics software.

Skewness is a measure of a distribution's deviation from symmetry. In a symmetrical distribution, the mean, median, and mode are in the same location. A distribution that has cases stretching toward one tail or the other is called *skewed.* As shown in Exhibit 16a-3, part B, when the tail stretches to the right, to larger values, it is positively skewed. In part C, scores stretching toward the left, toward smaller values, skew the distribution negatively. Note the relationship between the mean, median, and mode in asymmetrical distributions. The symbol for skewness is *sk*.

$$sk = \frac{n}{(n-1)(n-2)} \sum \left(\frac{x_i - \bar{x}}{s} \right)^3$$

where s is the sample standard deviation (the unbiased estimate of sigma).

When a distribution approaches symmetry, *sk* is approximately 0. With a positive skew, *sk* will be a positive number; with negative skew, *sk* will be a negative number. The calculation of skewness for our annual percentage unit sales increase data produces an index of 0 and reveals no skew.

As illustrated in the lower portion of Exhibit 16a-3, *kurtosis* is a measure of a distribution's peakedness (or flatness). Distributions that have scores that cluster heavily or pile up in the center (along with more observations than normal in the extreme tails) are peaked or *leptokurtic.* Flat distributions, with scores more evenly distributed and tails fatter than a normal distribution, are called *playkurtic.* Intermediate or *mesokurtic* distributions approach normal—neither too peaked nor too flat. The symbol for kurtosis is *ku.*

$$ku = \left[\frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum \left(\frac{x_i - \bar{x}}{s} \right)^4 \right] - \frac{3(n-1)^2}{(n-2)(n-3)}$$

where s is the sample standard deviation (the unbiased estimate of sigma).

The value of *ku* for a normal or mesokurtic distribution is close to 0. A leptokurtic distribution will have a positive value, and the playkurtic distribution will be negative. As with skewness, the larger the absolute value of the index, the more extreme is the characteristic. In the annual percentage unit sales increase example, the kurtosis is calculated as −.29, which suggested a very slight deviation from a normally shaped curve with some flattening contributed by smaller-than-expected frequencies of the value 7 in the example distribution.

>chapter 17

Exploring, Displaying, and Examining Data

❝It is precisely because the unexpected jolts us out of our preconceived notions, our assumptions, our certainties, that it is such a fertile source of innovation.❞

Peter Drucker, author, Innovation and Entrepreneurship

>learningobjectives

After reading this chapter, you should understand . . .

1 That exploratory data analysis techniques provide insights and data diagnostics by emphasizing visual representations of the data.

2 How cross-tabulation is used to examine relationships involving categorical variables, serves as a framework for later statistical testing, and makes table-based analysis using one or more control variables an efficient tool for data visualization and decision making.

"Sammye Grayson, meet Myra Wines from MindWriter. We'll be working with her on a short-turnaround project during the next week." Sammye rises to shake Myra's extended hand, as Jason asks, innocently, "Anything interesting on those initial cross-tabs?"

Myra smiles, raises an expressive eyebrow, and waits for Sammye's response.

Sammye hesitates and then, looking at Jason for some signal of why he asked the question in the presence of a different client, responds, "Three of the early cross-tabs appeared to show some support for the board's assumptions about the alcohol issue—on whether current patrons endorse the selling of beer and wine during intermissions. But we're not far enough into the data to say which of the board's assumptions are fully correct and which might have to be modified based on the patterns emerging within subgroups of the sample."

Jason raises a hand to stop the detailed answer to his question. Sammye knows from the look on his face that she's done something wrong.

"I shouldn't have answered your question," blurts Sammye. "I walked right into the trap you set, eyes wide open."

Myra jumps in before Jason can respond. "I've seen Jason do this once before to an intern, so you should feel like one of the team. And, no, you shouldn't have responded—confidentiality is rule number one—and as a client, I appreciate it. No harm done this time, though. What Jason failed to tell you is I'm on CCPA's board and part of the project team. Before Jason stopped you, things were getting interesting. Please continue."

Sammye, getting a nod from Jason, shares, "We'll probably have to do some recoding of the age and race variables for the patterns to emerge clearly. The team is also interested in the differences between ethnic groups in future performance preferences. We've also finished coding each patron's address with its GPS (Geographic Positioning System) code. The preliminary mapping begins tomorrow; Jason hired a master's candidate in geography to provide the mapping. I've scheduled a conference call for . . . (Sammye flips her desk calendar pages to the following week) . . . Friday of next week with Jackson Murray and other members of the CCPA project team."

"When the board approved your proposed analysis plan," queries Myra, "I don't remember seeing any reference to those boxlike diagrams with tails I see on that graph you just handed to Jason."

"Most of what the team will be doing the next three days," intervenes Jason, "involves more graphical displays than statistical ones. Right now we're just getting a sense of what the data are telling us. We'll decide what, if any, new analyses to add to the proposed plan by this Friday. It's this early work that lays the groundwork for the more sophisticated analyses that follow. There isn't anything glamorous about it, but without it we might miss some crucial findings."

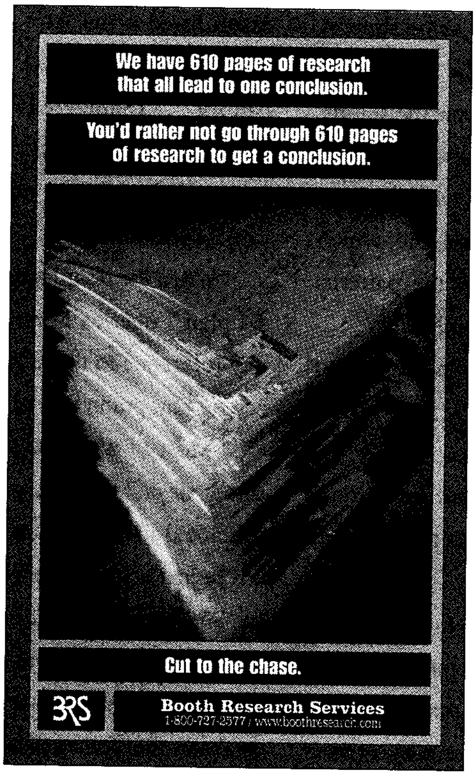
Jason pauses for effect and then says, "By the way, that 'little diagram' is called a boxplot. I actually did several during the preliminary analysis phase for MindWriter's CompleteCare study. I didn't give them to you because I would have had to explain how to interpret them and . . ."

". . . and anything you have to explain isn't clear enough," finishes Myra.

> Exploratory Data Analysis

The convenience of data entry via spreadsheet, optimal mark recognition (OMR), or the data editor of a statistical program makes it tempting to move directly to statistical analysis. That temptation is even stronger when the data can be entered and viewed in real time. Why waste time finding out if the data confirm the hypothesis that motivated the study? Why not obtain descriptive statistical summaries (based on our discussion in Appendix 16a) and then test hypotheses?

Exploratory data analysis is both a data analysis perspective and a set of techniques. In this chapter, we will present unique and conventional techniques including graphical and tabular devices to visualize the data. Exhibit 17-1 reminds you of the importance of data visualization as an integral element in the data analysis process and as a necessary step prior to hypothesis testing. In Chapter 2, we said research conducted scientifically is a puzzle-solving activity as well as an attitude of curiosity, suspicion, and imagination essential to discovery. It is natural, then, that exploration and examination of the data would be an integral part of our data analysis perspective.

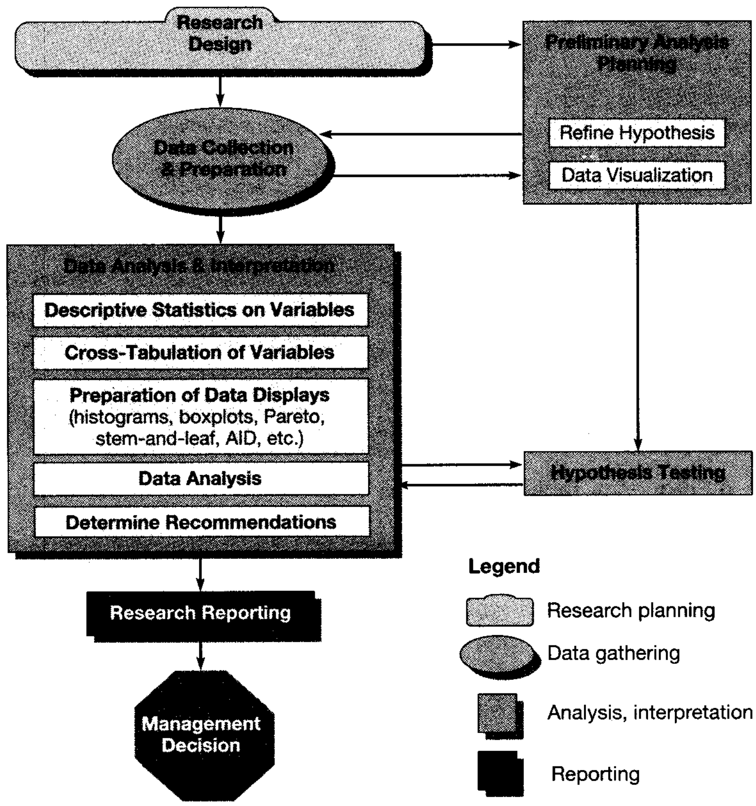


As this Booth Research Services ad suggests, the researcher's role is to make sense of numerous data displays and thus assist the research sponsor in making an appropriate decision. Great data exploration and analysis will distill mountains of data printouts into insightful and supportable conclusions. **www.boothresearch.com**

In **exploratory data analysis (EDA)** the researcher has the flexibility to respond to the patterns revealed in the preliminary analysis of the data. Thus patterns in the collected data guide the data analysis or suggest revisions to the preliminary data analysis plan. This flexibility is an important attribute of this approach. When the researcher is attempting to prove causation, however, confirmatory data analysis is required. **Confirmatory data analysis** is an analytical process guided by classical statistical inference in its use of significance testing and confidence.[1]

One authority has compared exporatory data analysis to the role of police detectives and other investigators and confirmatory analysis to that of judges and the judicial system. The former are involved in the search for clues and evidence; the latter are preoccupied with evaluating the strength of the evidence that is found. Exploratory data analysis is the first step in the search for

> Exhibit 17-1 Data Exploration, Examination, and Analysis in the Research Process



evidence, without which confirmatory analysis has nothing to evaluate.[2] Consistent with that analogy, EDA shares a commonality with exploratory designs, not formalized ones. Because it doesn't follow a rigid structure, it is free to take many paths in unraveling the mysteries in the data—to sift the unpredictable from the predictable.

A major contribution of the exploratory approach lies in the emphasis on visual representations and graphical techniques over summary statistics. Summary statistics, as you will see momentarily, may obscure, conceal, or even misrepresent the underlying structure of the data. When numerical summaries are used exclusively and accepted without visual inspection, the selection of confirmatory models may be based on flawed assumptions.[3] For these reasons, data analysis should begin with visual inspection. After that, it is not only possible but also desirable to cycle between exploratory and confirmatory approaches.

< Location, spread, and shape were introduced in Appexdix 16a.

Frequency Tables, Bar Charts, and Pie Charts[4]

Several useful techniques for displaying data are not new to EDA. They are essential to any examination of the data. For example, a **frequency table** is a simple device for arraying data. An example is presented in Exhibit 17-2. It arrays data by assigned numerical value, with columns for percent, valid percent (percent adjusted for missing data), and cumulative percent. Ad recall, a nominal variable, describes the ads that participants

>snapshot

The Media Outlook

Demand and economic forecasting are research staples for a marketer. Some firms like Wilkofsky Gruen (W&G) have built a business around developing such forecasts. Others, like auditor PricewaterhouseCoopers (PWC), crafted their reputation as an industry expert by underwriting such forecasts. These two firms annually collaborate to develop the entertainment industry's *Global Entertainment and Media Outlook*. "W&G has developed a set of proprietary algorithms, which use information from a variety of sources—a special PWC data collection project in a particular country, government databases, and their own proprietary databases, for example," shared Deborah Scruby, marketing director at PWC. "Once their draft report is prepared—it can be as large as 500 pages—we circulate it to experts and specialists around the world, who help interpret or find additional insights on various aspects of specific media and the entertainment industry." Together W&G and PWC collaborate to develop the final forecast, which is sold to financial analysts, large multinational media companies, universities and libraries, and the various media.

www.pwcglobal.com; www.wilkofskygruen.com

> Exhibit 17-2 A Frequency Table of Ad Recall

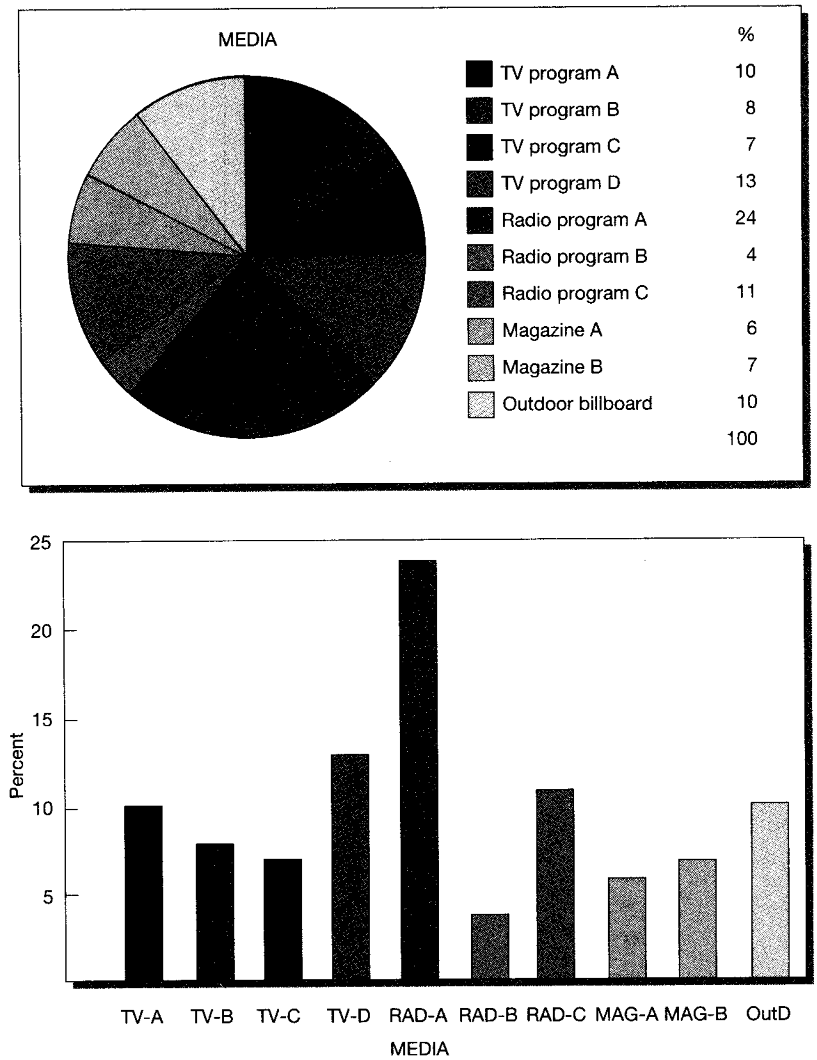
Value Label	Value	Frequency	Percent	Valid Percent	Cumulative Percent
TV program A	1	10	10.0	10.0	10.0
TV program B	2	8	8.0	8.0	18.0
TV program C	3	7	7.0	7.0	25.0
TV program D	4	13	13.0	13.0	38.0
Radio program A	5	24	24.0	24.0	62.0
Radio program B	6	4	4.0	4.0	66.0
Radio program C	7	11	11.0	11.0	77.0
Magazine A	8	6	6.0	6.0	83.0
Magazine B	9	7	7.0	7.0	90.0
Outdoor billboard	10	10	10.0	10.0	100.0
Total		100	100.0	100.0	

Valid cases 100 Missing cases 0

remembered seeing or hearing without being prompted by the researcher or the measurement instrument. Although there are 100 observations, the small number of media placements make the variable easily tabled. The same data are presented in Exhibit 17-3 using a pie chart and a bar chart. The values and percentages are more readily understood in this graphic format, and visualization of the media placements and their relative sizes is improved.

When the variable of interest is measured on an interval-ratio scale and is one with many potential values, these techniques are not particularly informative. Exhibit 17-4 is a condensed frequency table of the average annual purchases of PrimeSell's top 50 customers. Only two values, 59.9 and 66, have a frequency greater than 1. Thus, the primary contribution of this table is an ordered list of values. If the table were converted to a bar chart, it would have 48 bars of equal length and two bars with two occurrences. Bar charts do not reserve spaces for values where no observations occur within the range. Constructing a pie chart for this variable would also be pointless.

> **Exhibit 17-3** Nominal Variable Displays (Ad Recall)



Histograms

The histogram is a conventional solution for the display of interval-ratio data. **Histograms** are used when it is possible to group the variable's values into intervals. Histograms are constructed with bars (or asterisks) that represent data values, where each value occupies an equal amount of area within the enclosed area. Data analysts find histograms useful for (1) displaying all intervals in a distribution, even those without observed values, and (2) examining the shape of the distribution for skewness, kurtosis, and the modal pattern. When looking at a histogram, one might ask: Is there a single hump (a mode)? Are subgroups identifiable when multiple modes are present? Are straggling data values detached from the central concentration?[5]

The values for the average annual purchases variable presented in Exhibit 17-4 were measured on a ratio scale and are easily grouped. Other variables possessing an underlying order are similarly appropriate for

> **Exhibit 17-4** Average Annual Purchases of PrimeSell's Top 50 Customers

Value	Frequency	Percent	Cumulative Percent	Value	Frequency	Percent	Cumulative Percent
54.9	1	2	2	75.6	1	2	54
55.4	1	2	4	76.4	1	2	56
55.6	1	2	6	77.5	1	2	58
56.4	1	2	8	78.9	1	2	60
56.8	1	2	10	80.9	1	2	62
56.9	1	2	12	82.2	1	2	64
57.8	1	2	14	82.5	1	2	66
58.1	1	2	16	86.4	1	2	68
58.2	1	2	18	88.3	1	2	70
58.3	1	2	20	102.5	1	2	72
58.5	1	2	22	104.1	1	2	74
59.9	2	4	26	110.4	1	2	76
61.5	1	2	28	111.9	1	2	78
62.6	1	2	30	118.6	1	2	80
64.8	1	2	32	123.8	1	2	82
66.0	2	4	36	131.2	1	2	84
66.3	1	2	38	140.9	1	2	86
67.6	1	2	40	146.2	1	2	88
69.1	1	2	42	153.2	1	2	90
69.2	1	2	44	163.2	1	2	92
70.5	1	2	46	166.7	1	2	94
72.7	1	2	48	183.2	1	2	96
72.9	1	2	50	206.9	1	2	98
73.5	1	2	52	218.2	1	2	100
				Total	50	100	

histograms. A histogram would not be used for a nominal variable like ad recall (Exhibit 17-3) that has no order to its categories.

A histogram of the average annual purchases is shown in Exhibit 17-5. The midpoint for each interval for the variable of interest, average annual purchases, is shown on the horizontal axis; the frequency or number of observations in each interval, on the vertical axis. We erect a vertical bar above the midpoint of each interval on the horizontal scale. The height of the bar corresponds with the frequency of observations in the interval above which it is erected. This histogram was constructed with intervals 20 increments wide, and the last interval contains only two observations, 206.9 and 218.2. These values are found in PrimeSell's average annual purchases frequency table (Exhibit 17-4). Intervals with 0 counts show gaps in the data and alert the analyst to look for problems with spread. When the upper tail of the distribution is compared with the frequency table, we find three extreme values (183.2, 206.9, and 218.2). Along with the peaked midpoint and reduced number of observations in the upper tail, this histogram warns us of irregularities in the data.

>snapshot

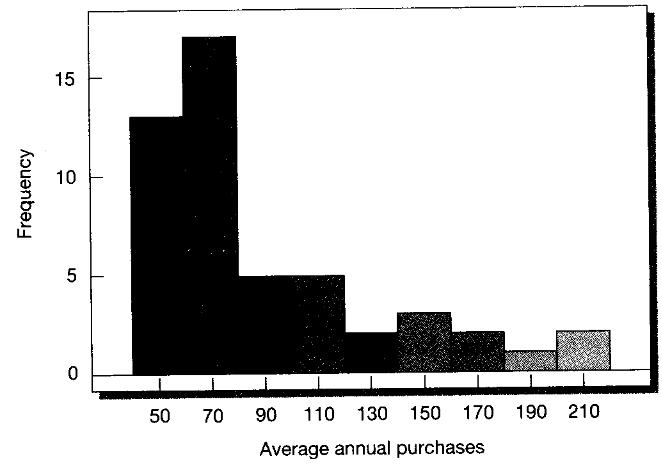
Wirthlin Worldwide Research Redesigns Red Cross Donations

September 11, 2001, elicited generous donations to the American Red Cross (RC) while it simultaneously raised questions and concerns about the Red Cross's fund-raising practices and processes, especially those related to the Liberty Fund. Realizing that confusion could erode trust, the RC hired Wirthlin Worldwide to assist it to improve its fund-raising messaging. Wirthlin conducted 12 focus groups in seven cities (Baltimore, Birmingham, Boston, Charleston, Houston, Minneapolis, and San Francisco)— to represent a range of urban and suburban populations, different degrees of local disasters and diversity, and city population sizes—followed by two nationwide public opinion polls. Two key segments were of interest in both surveys, each involving 1,000 adults: (1) prior Red Cross donors and (2) those who had never donated to the Red Cross. Survey participants were asked to listen to 30-second radio ads and asked questions that would test the clarity and persuasiveness of the revised messaging that evolved out of the focus groups. Following the survey, the RC consulted nonprofit opinion leaders (e.g., Better Business Bureau, GuideStar, Association of Fundraising Professionals) to evaluate the proposed fund-raising changes that were indicated by the survey. A new fund-raising process—Donor DIRECT (Donor Intent Recognition Confirmation Trust)—was launched June 2002, receiving favorable publicity and positive subsequent polling feedback. The Colorado fires during the summer of 2002 provided the litmus test of the new process: the Red Cross saw an increase in undesignated donations to its Disaster Relief Fund. The Red Cross and Wirthlin Worldwide research earned *PRWeek's* 2003 Technique Award: Best Use of Research or Measurement.

www.wirthlinworldwide.com; www.redcross.org

To learn more about this research, read the case: "Can Research Rescue the Red Cross?"

> Exhibit 17-5 Histogram of PrimeSell's Top 50 Customers' Average Annual Purchases



Stem-and-Leaf Displays[6]

The **stem-and-leaf display** is a technique that is closely related to the histogram. It shares some of the histogram's features but offers several unique advantages. It is easy to construct by hand for small samples or may be produced by computer programs. In contrast to histograms, which lose information by grouping data values into intervals, the stem-and-leaf presents actual data values that can be inspected directly, without the use of enclosed bars or asterisks as the representation medium. This feature reveals the distribution of values

> Exhibit 17-6 A Stem-and-Leaf Display of PrimeSell's Average Annual Purchases Data

5	4 5 5 6 6 6 7 8 8 8 8 9
6	1 2 4 6 6 7 9 9
7	0 2 2 3 5 6 7 8
8	0 2 2 6 8
9	
10	2 4
11	0 1 8
12	3
13	1
14	0 6
15	3
16	3 6
17	
18	3
19	
20	6
21	8

within the interval and preserves their rank order for finding the median, quartiles, and other summary statistics. It also eases linking a specific observation back to the data file and to the subject that produced it.

Visualization is the second advantage of stem-and-leaf displays. The range of values is apparent at a glance, and both shape and spread impressions are immediate. Patterns in the data—such as gaps where no values exist, areas where values are clustered, or outlying values that differ from the main body of the data—are easily observed.

To develop a stem-and-leaf display for the data in Exhibit 17-4, the first digits of each data item are arranged to the left of a vertical line. Next, we pass through the average annual purchases percentages in the order they were recorded and place the last digit for each item (the unit position, 1.0) to the right of the vertical line. Note that the digit to the right of the decimal point is ignored. The last digit for each item is placed on the horizontal row corresponding to its first digit(s). Now it is a simple matter to rank-order the digits in each row, creating the stem-and-leaf display shown in Exhibit 17-6.

Each line or row in this display is referred to as a *stem,* and each piece of information on the stem is called a *leaf.* The first line or row is

5 | 4 5 5 6 6 6 7 8 8 8 8 9

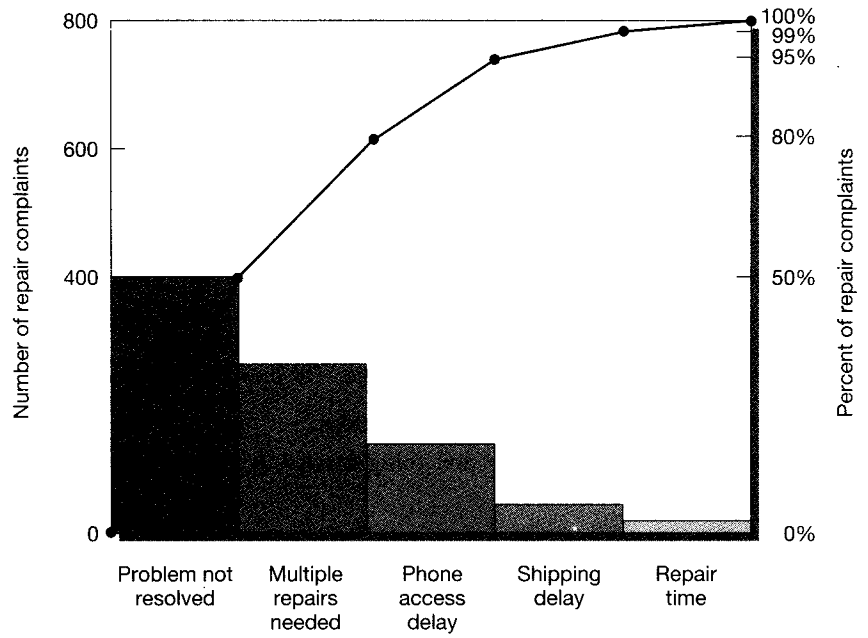
The meaning attached to this line or row is that there are 12 items in the data set whose first digit is five: 54, 55, 55, 56, 56, 56, 57, 58, 58, 58, 58, and 59. The second line,

6 | 1 2 4 6 6 7 9 9

shows that there are eight average annual purchase values whose first digit is six: 61, 62, 64, 66, 66, 67, 69, and 69.

When the stem-and-leaf display shown in Exhibit 17-6 is turned upright (rotated 90 degrees to the left), the shape is the same as that of the histogram shown in Exhibit 17-5.

> **Exhibit 17-7** Pareto Diagram of MindWriter Repair Complaints



Pareto Diagrams

Pareto diagrams derive their name from a 19th-century Italian economist. In quality management, J. M. Juran first applied this concept by noting that only a vital few defects account for most problems evaluated for quality and that the trivial may explain the rest. Historically, this has come to be known as the 80/20 rule—that is, an 80 percent improvement in quality or performance can be expected by eliminating 20 percent of the causes of unacceptable quality or performance.

The **Pareto diagram** is a bar chart whose percentages sum to 100 percent. The data are derived from a multiple-choice, single-response scale; a multiple-choice, multiple-response scale; or frequency counts of words (or themes) from content analysis. The respondents' answers are sorted in decreasing importance, with bar height in descending order from left to right. The pictorial array that results reveals the highest concentration of improvement potential in the fewest number of remedies. An analysis of MindWriter customer complaints is depicted as a Pareto diagram in Exhibit 17-7. The cumulative frequency line in this exhibit shows that the top two problems (the repair did not resolve the customer's problem, and the product was returned multiple times for repair) accounted for 80 percent of the perceptions of inadequate repair service.

Boxplots[7]

The **boxplot,** or *box-and-whisker plot,* is another technique used frequently in exploratory data analysis.[8] A boxplot reduces the detail of the stem-and-leaf display and provides a different visual image of the distribution's location, spread, shape, tail length, and outliers. Boxplots are extensions of the **five-number summary** of a distribution. This summary consists of the median, the upper and lower quartiles, and the largest and smallest observations. The median and quartiles are used because they are particularly **resistant statistics.** *Resistance* is a characteristic that "provides insensitivity to localized misbehavior in data."[9] Resistant statistics are unaffected by outliers and change only slightly in response to the replacement of small portions of the data set.

Recall the discussion of the mean and standard deviation in Appendix 16a. Now assume we take a data set [5,6,6,7,7,7,8,8,9] and calculate its mean. The mean of the set is 7; the standard deviation 1.22. If the 9 is replaced with 90, the mean becomes 16 and the standard deviation increases to 27.78. The mean is now 2 times larger than most of the numbers in the distribution, and the standard deviation is more than 22 times its original size. Changing only one of nine values has disturbed the location and spread summaries to the point where they no longer represent the other eight values. Both the mean and the standard deviation are considered **nonresistant statistics;** they are susceptible to the effects of extreme values in the tails of the distribution and do not represent typical values well under conditions of asymmetry. The standard deviation is particularly problematic because it is computed from the squared deviations from the mean.[10] In contrast, the median and quartiles are highly resistant to change. When we changed the 9 to 90, the median remained at 7 and the lower and upper quartiles stayed at 6 and 8, respectively. Because of the nature of quartiles, up to 25 percent of the data can be made extreme without perturbing the median, the rectangular composition of the plot, or the quartiles themselves. These characteristics of resistance are incorporated into the construction of boxplots.

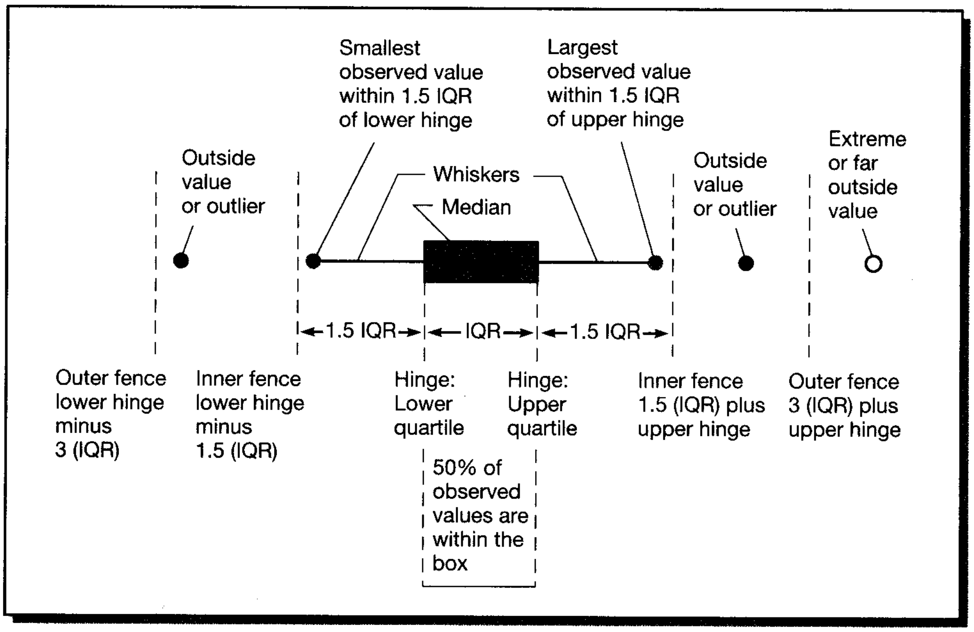
Boxplots may be constructed easily by hand or by computer programs. The basic ingredients of the plot are:

1. The rectangular plot that encompasses 50 percent of the data values.
2. A center line (or other notation) marking the median and going through the width of the box.
3. The edges of the box, called *hinges.*
4. The "whiskers" that extend from the right and left hinges to the largest and smallest values.[11]

These values may be found within 1.5 times the **interquartile range (IQR)** from either edge of the box. These components and their relationships are shown in Exhibit 17-8.

When you are examining data, it is important to separate legitimate outliers from errors in measurement, editing, coding, and data entry. **Outliers,** data points that exceed + 1.5 the interquartile range, reflect unusual cases and are an important source of information for the study. They are displayed or given special statistical

> Exhibit 17-8 Boxplot Components



treatment, or other portions of the data set are sometimes shielded from their effects. Outliers that are entry mistakes should be corrected or removed during editing.

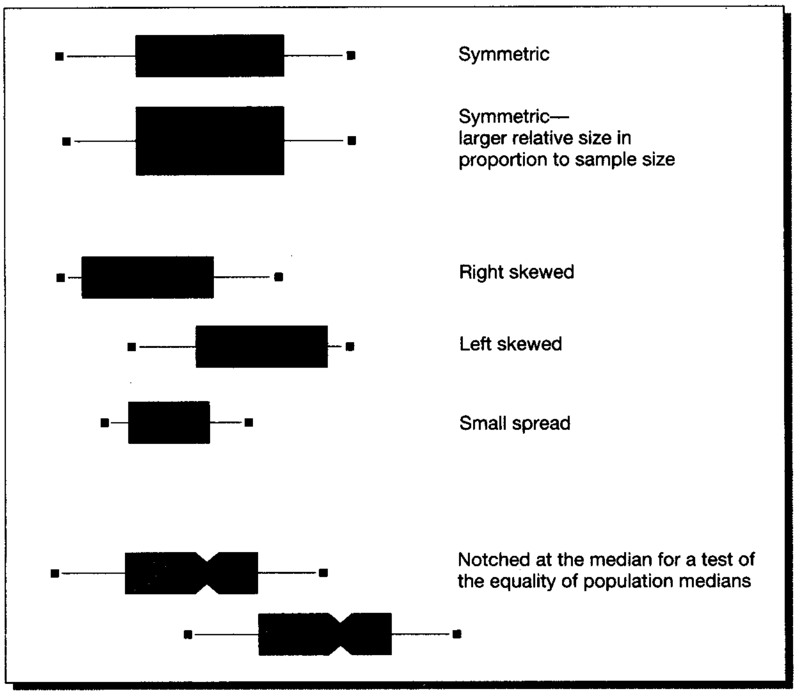
Exhibit 17-9 summarizes several comparisons that are of help to the analyst. Boxplots are an excellent diagnostic tool, especially when graphed on the same scale. The upper two plots in the exhibit are both symmetric, but one is larger than the other. Larger box widths are sometimes used when the second variable, from the same measurement scale, comes from a larger sample size. The box widths should be proportional to the square root of the sample size, but not all plotting programs account for this.[12] Right- and left-skewed distributions and those with reduced spread are also presented clearly in the plot comparison. Finally, groups may be compared by means of multiple plots. One variation, in which a notch at the median marks off a confidence interval to test the equality of group medians, takes us a step closer to hypothesis testing.[13] Here the sides of the box return to full width at the upper and lower confidence intervals. When the intervals do not overlap, we can be confident, at a specified confidence level, that the medians of the two populations are different.

In Exhibit 17-10, multiple boxplots compare five sectors of PrimeSell's customers by their average annual purchases data. The overall impression is one of potential problems for the analyst: unequal variances, skewness, and extreme outliers. Note the similarities of the profiles of finance and retailing in contrast to the high-tech and insurance sectors. If hypothesis tests are planned, further examination of this plot for each sector would require a stem-and-leaf display and a five-number summary. From this, we could make decisions on the types of tests to select for confirmatory analysis (see Chapters 18, 19, and 20).

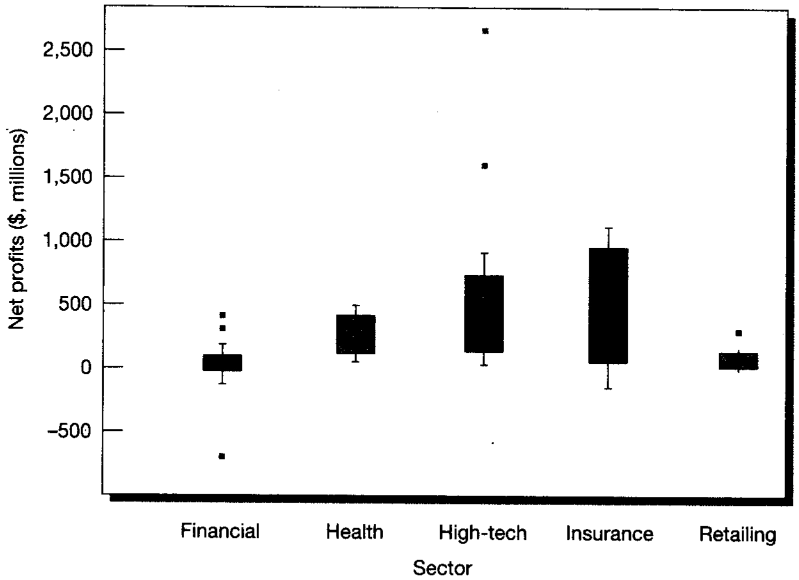
Mapping

Increasingly, participant data are being attached to their geographic dimension as Geographic Information System (GIS) software and coordinate measuring devices have become more affordable and easier to use. Essentially a GIS works by linking data sets to each other with at least one common data field (for example, a

> Exhibit 17-9 Diagnostics with Boxplots



> **Exhibit 17-10** Boxplot Comparison of Customer Sectors



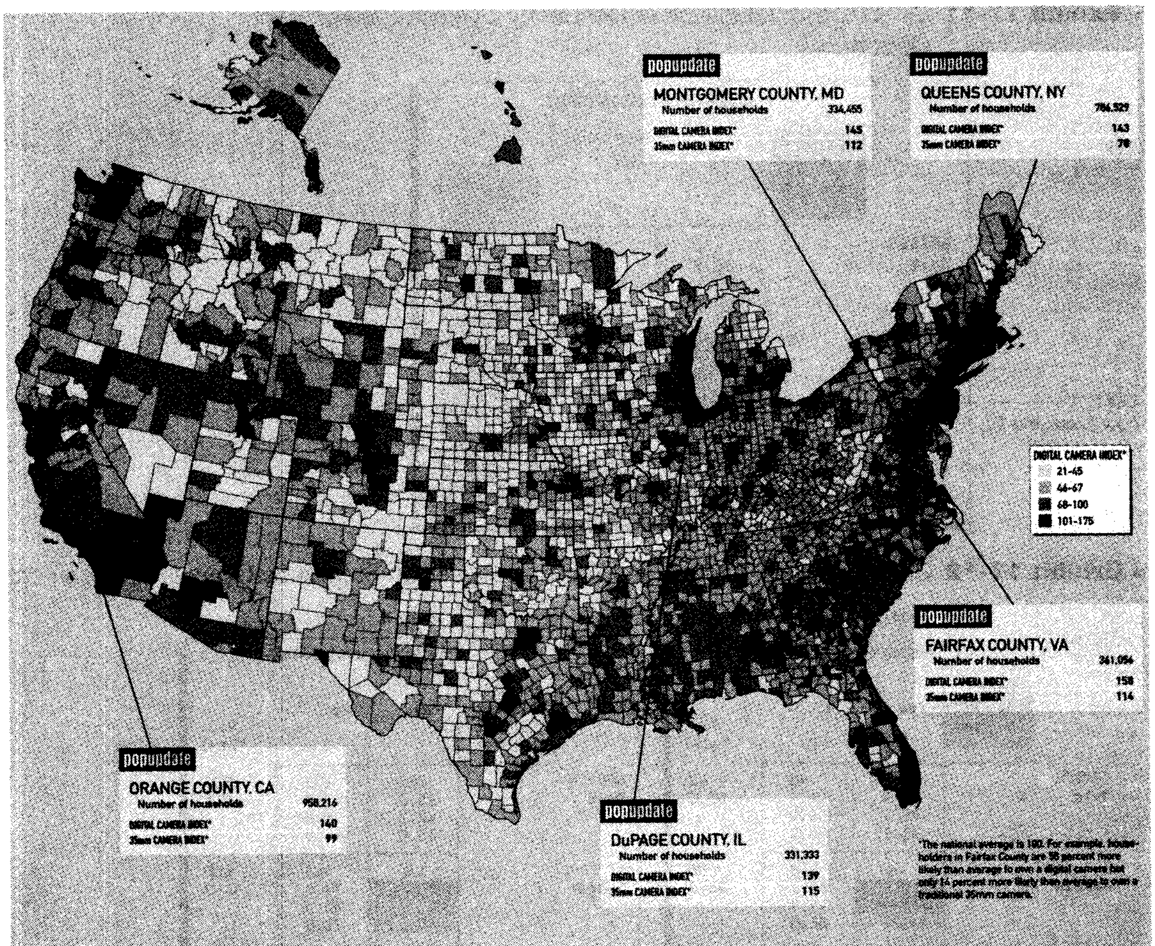
household's street address). The GIS allows the researcher to connect target and classification variables from a survey to specific geographic-based databases like U.S. Census data, to develop a richer understanding of the sample's attitudes and behavior. When radio frequency identification (RFID) data become more prevalent, much behavioral data will be able to connect with these new geographically rich databases.

The most common way to display such data is with a map. Colors and patterns denoting knowledge, attitude, behavior, or demographic data arrays are superimposed over street maps (finest-level GIS), block-group maps, or county, state, or country maps to help identify the best locations for stores based on demographic, psychographic, and life-stage segmentation data. Florists array promotional response information geographically and use the map to plan targeted promotions. Consumer and business-to-business researchers use mapping of data on ownership, usage level, and price sensitivity in plotting geographic rollouts of new products. While this is an attractive option for exploratory analysis, it does take specialized software and hardware, as well as the expertise to operate it. Students are encouraged to take specialized courses on GIS to expand their skill set in this growing area.

Throughout this section we have exploited the visual techniques of exploratory data analysis to look beyond numerical summaries and gain insight into the patterns of the data. Few of the approaches have stressed the need for advanced mathematics, and all have an intuitive appeal for the analyst. When the more common ways of summarizing location, spread, and shape have conveyed an inadequate picture of the data, we have used more resistant statistics to protect us from the effects of extreme scores and occasional errors. We have also emphasized the value of transforming the original scale of the data during preliminary analysis rather than at the point of hypothesis testing.

> Cross-Tabulation

Depending on the management question, we can gain valuable insights by examining the data with cross-tabulation. **Cross-tabulation** is a technique for comparing data from two or more categorical variables such as gender and selection by one's company for an overseas assignment. Cross-tabulation is used with



Sometimes there is no better way to display data than with a map. In 2003, sales of digital cameras exceeded $3.1 billion, more than 10 times the level in 1998. That's bad news for companies like Kodak that once enjoyed significant revenue from repeat purchases of film. To find counties where ownership of digital cameras is above average, *American Demographics* teamed up with Claritas to design this map. At a quick glance you can tell areas where digital cameras have really captured the shutterbug's fancy and areas where penetration is limited. Photo Marketing Association International estimates U.S. household penetration at 42 percent in 2004 and forecasts sales of 25.8 million digital cameras in 2008. **http://demographics.com; www.claritas.com; www.pmai.org.**

demographic variables and the study's target variables (operationalized measurement questions). The technique uses tables having rows and columns that correspond to the levels or code values of each variable's categories. Exhibit 17-11 is an example of a computer-generated cross-tabulation. This table has two rows for gender and two columns for assignment selection. The combination of the variables with their values produces four cells. Each **cell** contains a count of the cases of the joint classification and also the row, column, and total percentages. The number of row cells and column cells is often used to designate the size of the table, as in this 2 × 2 table. The cells are individually identified by their row and column numbers, as illustrated. Row and column totals, called **marginals,** appear at the bottom and right "margins" of the table. They show the counts and percentages of the separate rows and columns.

Cross-tabulation is a first step for identifying relationships between variables. When tables are constructed for statistical testing, we call them **contingency tables,** and the test determines if the classification variables are independent of each other (see chi-square in Chapter 18). Of course, tables may be larger than 2 × 2.

> Exhibit 17-11 SPSS Cross-Tabulation of Gender by Overseas Assignment Opportunity

Cell content

Cell 2, 1 (row 2, column 1)

Marginals

		OVERSEAS ASSIGNMENT		
	[redacted]	Yes 1	No 2	Row Total
GENDER Male	1	22 35.5 78.6 22.0	40 64.5 55.6 40.0	62 62.0
Female	2	[redacted]	32 84.2 44.4 32.0	38 38.0
	Column Total	28 28.0	72 72.0	100 100.0

> Exhibit 17-12 Comparison of Percentages in Cross-Tabulation Studies by Overseas Assignment

Study 1

		OVERSEAS ASSIGNMENT		
	Count [redacted] Tot Pct	Yes 1	No 2	Row Total
GENDER Male	1	22 35.5 78.6 22.0	40 64.5 55.6 40.0	62 62.0
Female	2	6 [redacted] 6.0	32 84.2 44.4 32.0	38 38.0
	Column Total	28 28.0	72 72.0	100 100.0

Study 2

		OVERSEAS ASSIGNMENT		
	Count Row Pct Col Pct Tot Pct	Yes 1	No 2	Row Total
GENDER Male	1	225 25.0 62.5 15.0	675 75.0 59.2 45.0	900 60.0
Female	2	135 [redacted] 9.0	465 77.5 40.8 31.0	600 40.0
	Column Total	360 24.0	1140 76.0	1500 100.0

The Use of Percentages

Percentages serve two purposes in data presentation. First, they simplify the data by reducing all numbers to a range from 0 to 100. Second, they translate the data into standard form, with a base of 100, for relative comparisons. In a sampling situation, the number of cases that fall into a category is meaningless unless it is related to some base. A count of 28 overseas assignees has little meaning unless we know it is from a sample of 100. Using the latter as a base, we conclude that 28 percent of this study's sample has an overseas assignment.

While the above is useful, it is even more useful when the research problem calls for a comparison of several distributions of data. Assume the previously reported data were collected five years ago and the present study had a sample of 1,500, of which 360 were selected for overseas assignments. By using percentages, we can see the relative relationships and shifts in the data (see Exhibit 17-12).

With two-dimension tables, the selection of a row or column will accentuate a particular distribution or comparison. This raises the question about which direction the percentages should be calculated. Most computer programs offer options for presenting percentages in both directions and interchanging the rows and

>snapshot

Extensive Research Launches Starbucks Card Duetto™ Visa

BusinessWeek recognized the Starbucks Card Duetto™ Visa as one of the important new products of 2003. In fact, it was the only financial product on the list. Starbucks Card Duetto™ Visa is a multifunction card that combines the features of a prepaid stored-value card, known as the Starbucks Card, with a regular credit card. Starbucks, in conjunction with Visa and Bank One, did extensive research to determine if the proposed new payment option had appeal. Focus groups were used to determine the level of potential consumer confusion with the multifunction card, determine card attractiveness, and refine messaging. A series of online surveys were conducted both before and after the launch of the product to determine market receptivity. A press release about the partnership about eight months before launch generated news coverage resulting in traffic to Starbucks' Web site. Early purchase intent was determined by those Starbucks customers who took initiative and signed up via the Web site to be prenotified by e-mail when the card became available. Among the postlaunch research questions guiding measurement of return on marketing investment (ROMI) are:

- Does the card enhance the Starbucks customer experience (how satisfied is each customer, and do customers feel appreciated)?
- Did the card prove valuable to all partners: Starbucks, Bank One, and Visa?
- Did card activity, which is linked to charitable donations, permit Starbucks to give back to the communities in which it operates in a significant way?

If you were in charge of this research, what would you be looking for during exploratory data analysis?

www.bankone.com; www.starbucks.com; www.visa.com

To learn more about this research, read the case and watch the video "Starbucks, Bank One, and Visa Launch the Starbucks Card Duetto™ Visa."

columns of the table. But in situations where one variable is hypothesized to be the presumed cause, is thought to affect or predict a response, or is simply antecedent to the other variable, we label it the independent variable. Percentages should then be computed in the direction of this variable. Thus, if the independent variable is placed on the row, select row percentages; if it is on the column, select column percentages. In which direction should the percentages run in the previous example? If only the column percentages are reported, we imply that assignment status has some effect on gender. This is implausible. When percentages are reported by rows, the implication is that gender influences selection for overseas assignments.

Care should be taken in interpreting percentages from tables. Consider again the data in Exhibit 17-12. From the first to the second study, it is apparent that the percentage of females selected for overseas assignments rose from 15.8 to 22.5 percent of their respective samples. This should not be confused with the percentage within each sample who were women with overseas assignments, a number which increased from 6 percent (Study 1) to 9 percent (Study 2). Among all overseas selectees, in the first study 21.4 percent were women, while in the second study, 37.5 percent were women. Similar comparisons can be made for the other categories. The tables verify an increase in women with overseas assignments, but we cannot conclude that their gender had anything to do with the increase.

Percentages are used by virtually everyone dealing with numbers—but often incorrectly. The following guidelines, if used during analysis, will help to prevent errors in reporting:[14]

- *Averaging percentages.* Percentages cannot be averaged unless each is weighted by the size of the group from which it is derived. Thus, a simple average will not suffice; it is necessary to use a weighted average.
- *Use of too large percentages.* This often defeats the purpose of percentages—which is to simplify. A large percentage is difficult to understand and is confusing. If a 1,000 percent increase is experienced, it is better to describe this as a 10-fold increase.

- *Using too small a base.* Percentages hide the base from which they have been computed. A figure of 60 percent when contrasted with 30 percent would appear to suggest a sizable difference. Yet if there are only three cases in the one category and six in the other, the differences would not be as significant as they have been made to appear with percentages.
- *Percentage decreases can never exceed 100 percent.* This is obvious, but this type of mistake occurs frequently. The higher figure should always be used as the base or denominator. For example, if a price was reduced from $1 to $.25, the decrease would be 75 percent (75/100).

Other Table-Based Analysis

The recognition of a meaningful relationship between variables generally signals a need for further investigation. Even if one finds a statistically significant relationship, the questions of why and under what conditions remain. The introduction of a **control variable** to interpret the relationship is often necessary. Cross-tabulation tables serve as the framework.

Statistical packages like Minitab, SAS, and SPSS have among their modules many options for the construction of *n*-way tables with provision for multiple control variables. Suppose you are interested in creating a cross-tabulation of two variables with one control. Whatever the number of values in the primary variables, the control variable with five values determines the number of tables. For some applications, it is appropriate to have five separate tables; for others, it might be preferable to have adjoining tables or have the values of all the variables in one. Management reports are of the latter variety. Exhibit 17-13 presents an example in which all three variables are handled under the same banner. Programs such as this one can handle far more complex tables and statistical information.[15]

An advanced variation on *n*-way tables is **automatic interaction detection (AID).** AID is a computerized statistical process that requires that the researcher identify a dependent variable and a set of predictors or

> Exhibit 17-13 SPSS Cross-Tabulation with Control and Nested Variables

	Control Variable					
	Category 1			Category 2		
	Nested Variable			Nested Variable		
	cat 1	cat 2	cat 3	cat 1	cat 2	cat 3
Stub...	Cells...					

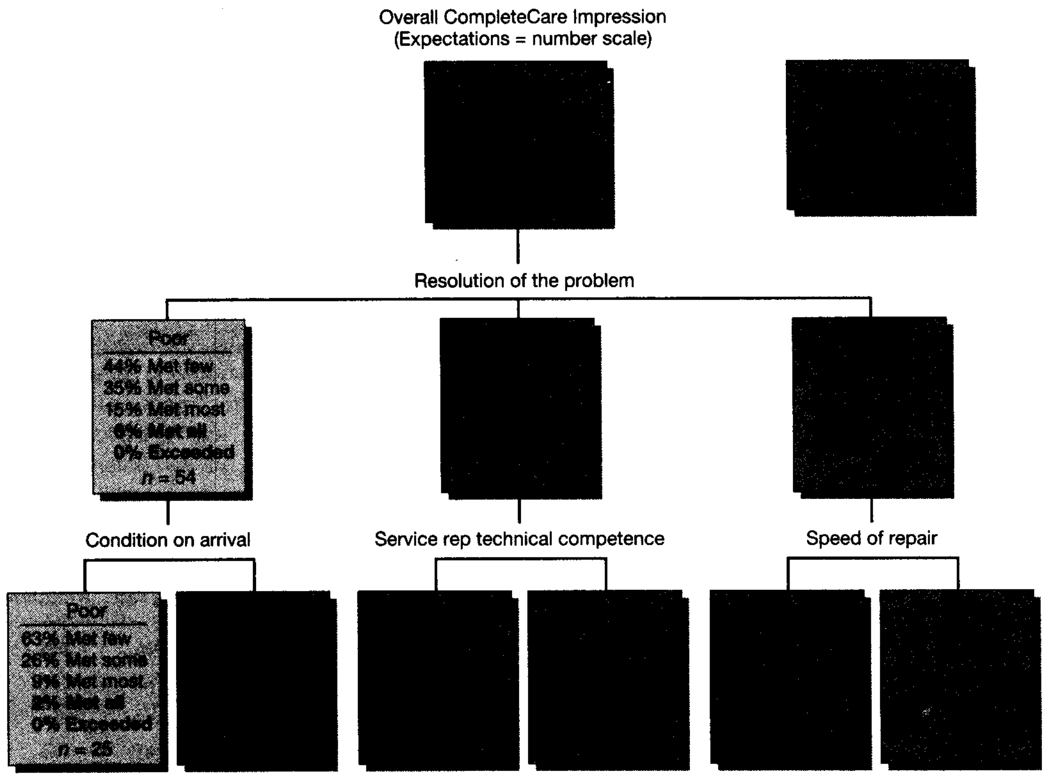
	SEX OF EMPLOYEE			
	MALES		FEMALES	
	MINORITY CLASSIFICATION		MINORITY CLASSIFICATION	
	WHITE	NONWHITE	WHITE	NONWHITE
EMPLOYMENT CATEGORY				
CLERICAL	16%	7%	18%	7%
OFFICE TRAINEE	7%	3%	17%	2%
SECURITY OFFICER	3%	3%		
COLLEGE TRAINEE	7%	0%	1%	
EXEMPT EMPLOYEE	6%	0%	0%	
MBA TRAINEE	1%	0%	0%	
TECHNICAL	1%			

independent variables. The computer then searches among up to 300 variables for the best single division of the data according to each predictor variable, chooses one, and splits the sample using a statistical test to verify the appropriateness of this choice.

Exhibit 17-14 shows the tree diagram that resulted from an AID study of customer satisfaction with MindWriter's CompleteCare repair service. The initial dependent variable is the overall impression of the repair service. This variable was measured on an interval scale of 1 to 5. The variables that contribute to perceptions of repair effectiveness were also measured on the same scale but were rescaled to ordinal data for this example (1–2 = poor, 3 = average, and 4–5 = excellent). The top box shows that 62 percent of the respondents rated the repair service as excellent (41% + 21%). The best predictor of repair effectiveness is "resolution of the problem."

On the left side of the tree, customers who rated "resolution of the problem" as poor have fewer expectations being met or exceeded than the average for the sample (6 percent versus 62 percent). A poor rating on "condition on arrival" exacerbates this, reducing the total satisfied group to 2 percent. From this example you can see that the researcher separately studied (applied AID to) each subgroup to find the variable that when split again makes the next largest contribution to understanding the consumers' evaluation process—and to the reduction of unexplained variation in each subsample. This analysis alerts decision makers at MindWriter to the best- and worst-case scenarios for the CompleteCare service, how to recover during a problematic month, and which "key drivers," or independent variables influencing the process, should receive corrective resources.

> Exhibit 17-14 Automatic Interaction Detection Example (MindWriter's Repair Satisfaction)



>summary

1 Exploratory data analysis (EDA) provides a perspective and set of tools to search for clues and patterns in the data. EDA augments rather than supplants traditional statistics. In addition to numerical summaries of location, spread, and shape, EDA uses visual displays to provide a complete and accurate impression of distributions and variable relationships.

Frequency tables array data from lowest to highest values with counts and percentages. They are most useful for inspecting the range of responses and their repeated occurrence. Bar charts and pie charts are appropriate for relative comparisons of nominal data. Histograms are optimally used with continuous variables where intervals group the responses. The Pareto diagram is a bar chart whose percentages sum to 100 percent. The causes of the problem under investigation are sorted in decreasing importance, with bar height descending from left to right. Stem-and-leaf displays and boxplots are EDA techniques that provide visual representations of distributions. The former present actual data values using a histogram-type device that allows inspection of spread and shape. Boxplots use the five-number summary to convey a detailed picture of a distribution's main body, tails, and outliers. Both stem-and-leaf displays and boxplots rely on resistant statistics to overcome the limitations of descriptive measures that are subject to extreme scores.

2 The examination of relationships involving categorical variables employs cross-tabulation. The tables used for this purpose consist of cells and marginals. The cells may contain combinations of count, row, column, and total percentages. The tabular structure is the framework for later statistical testing. Computer software for cross-classification analysis makes table-based analysis with one or more control variables an efficient tool for data visualization and later decision making. An advanced variation on n-way tables is automatic interaction detection (AID).

>keyterms

automatic interaction detection (AID) 486
boxplot 479
cell 483
confirmatory data analysis 472
contingency table 483
control variable 486
cross-tabulation 482
exploratory data analysis (EDA) 472
five-number summary 479
frequency table 473
histogram 475
interquartile range (IQR) 480
marginals 483
nonresistant statistics 480
outliers 480
Pareto diagram 479
resistant statistics 479
stem-and-leaf display 477

>discussionquestions

Terms in Review

1 Define or explain:
   a Marginals.
   b Pareto diagram.
   c Nonresistant statistics.
   d Lower control limit.
   e The five-number summary.

Making Research Decisions

2 Suppose you were preparing two-way tables of percentages for the following pairs of variables. How would you run the percentages?
   a Age and consumption of breakfast cereal.
   b Family income and confidence about the family's future.
   c Marital status and sports participation.
   d Crime rate and unemployment rate.

3 You study the attrition of entering college freshmen (those students who enter college as freshmen but don't stay to graduate). You find the following relationships between attrition, aid, and distance of home from college. What is your interpretation? Consider all variables and relationships.

	Aid		Home Near Receiving Aid		Home Far Receiving Aid	
	Yes (%)	No (%)	Yes (%)	No (%)	Yes (%)	No (%)
Drop Out	25	20	5	15	30	40
Stay	75	80	95	85	70	60

4 A local health agency is experimenting with two appeal letters, A and B, with which to raise funds. It sends out 400 of the A appeal and 400 of the B appeal (each subsample is divided equally among working-class and middle-class neighborhoods). The agency secures the results shown in the following table.

a Which appeal is the best?

b Which class responded better to which letter?

c Is appeal or social class a more powerful independent variable?

	Appeal A		Appeal B	
	Middle Class (%)	**Working Class (%)**	**Middle Class (%)**	**Working Class (%)**
Contribution	20	40	15	30
No Contribution	80	60	85	70
	100	100	100	100

5 Assume you have collected data on sales associates of a large retail organization in a major metropolitan area. You analyze the data by type of work classification, education level, and whether the workers were raised in a rural or urban setting. The results are shown below. How would you interpret them?

Annual Retail Employee Turnover per 100 Employees

	High Education		Low Education			
	Salaried	**Hourly Wage**	**Salaried**	**Hourly Wage**	**Salaried**	**Hourly Wage**
Rural	8	16	6	14	18	18
Urban	12	16	10	12	19	20

Bringing Research to Life

6 Identify the variables being cross-tabulated by Sammye. Identify some plausible reasons why such an exploration would be a good idea.

From Concept to Practice

7 Use the data in Exhibit 17-5 to construct a stem-and-leaf display.

a Where do you find the main body of the distribution?

b How many values reside outside the inner fence(s)?

>wwwexercise

Your university likely has SPSS or SAS, or SPSS came bundled with your book. Visit these two companies' Web sites and compare their software's statistical analysis capabilities. If you were buying such software for your employer, which would you choose and why?

>cases*

Agri Comp

NCRCC: Teeing Up and New Strategic Direction

Mastering Teacher Leadership

* All cases appear on the text CD; you will find abstracts of these cases in the Case Abstracts section of this text.

>chapter 18

Hypothesis Testing

❛❛People are 'erroneously confident' in their knowledge and underestimate the odds that their information or beliefs will be proved wrong. They tend to seek additional information in ways that confirm what they already believe.❜❜

Max Bazerman, Harvard University

>**learning**objectives

After reading this chapter, you should understand . . .

1 The nature and logic of hypothesis testing.

2 What a statistically significant difference is.

3 The six-step hypothesis testing procedure.

4 The differences between parametric and nonparametric tests and when to use each.

5 The factors that influence the selection of an appropriate test of statistical significance.

6 How to interpret the various test statistics.

"Sally, I'd like to meet with you about verifying the gender and age differences on the alcohol issue for Center City for Performing Arts Association." Jason makes his way through the cluttered outer office, stepping around piles of printouts, topped with sketched graphs or detailed cross-tabulated tables with handwritten notes.

Moments later, Sally enters, carrying the cross-tabulated data to which Jason had referred.

Jason smiles, waiting for her to settle. "So what have you got?"

"There is definitely a difference in attitude about serving alcohol during intermission at performances. But it doesn't appear to be quite what the CCPA board expected."

"How so?"

"Well, the younger patrons seem somewhat divided, while those between 35 and 54 are against and those 55 and over are strongly in favor."

"What was your original hypothesis?"

"Based on your meeting notes from the project session with the CCPA board, I formulated the hypothesis that there would be a difference on the alcohol issue based on age," shares Sally. "But I assumed that the younger the patrons, the more in favor of alcohol they would be. The numbers just aren't supporting that. And I'm not so sure that age is the right variable to look at."

Jason extends his hand across his desk. "Let me see the statistics on age."

"And I've got the stats on gender too," offers Sally.

"Are those in line with your hypothesis?"

"Not really," shares Sally. "Men and women are all over the place. I hypothesized that men would be in favor while women would be against. That's just not panning out."

Jason glances at the printout, pleased to see that her interpretation of the statistics is correct. "Looks like you have some work yet, to determine the pockets of resistance. Since the sample split—wasn't it 57 percent in favor to 43 percent against?—we don't want to recommend that CCPA proceed *without* being able to tell the board the likely direction of potential trouble.

"Sometimes our preliminary analysis plan can take us only so far," comments Jason. "Let's talk about what tests you plan to run now."

> Introduction

< **Induction and deduction were discussed in Chapter 2.**

In Chapters 16 and 17, we discussed the procedures for data preparation and preliminary analysis. The next step for many studies is hypothesis testing.

Just as your understanding of scientific reasoning was important in the last two chapters, recollection of the specific differences between induction and deduction is fundamental to hypothesis testing. Inductive reasoning moves from specific facts to general, but tentative, conclusions. We can never be absolutely sure that inductive conclusions are flawless. With the aid of probability estimates, we can qualify our results and state the degree of confidence we have in them. Statistical inference is an application of inductive reasoning. It allows us to reason from evidence found in the sample to conclusions we wish to make about the population.

> **A section on nonparametric techniques in Appendix B provides further study for readers with a special interest in nominal and ordinal variables.**

Inferential statistics is the second of two major categories of statistical procedures, the other being descriptive statistics. We used descriptive statistics in Chapter 15 when describing distributions. Under the heading **inferential statistics,** two topics are discussed in this book. The first, estimation of population values, was used with sampling in Chapter 15, but we will return to it here briefly. The second, testing statistical hypotheses, is the primary subject of this chapter. There are more examples of hypothesis tests in this chapter than most students will need for a term project or early assignments in their research careers. They are provided, like Appendix B, for later reference along with the readings for this chapter found at the end of the book.

Having detailed your hypotheses in your preliminary analysis planning, the purpose of hypothesis testing is to determine the accuracy of your hypotheses due to the fact that you have collected a sample of data, not a census. Exhibit 18-1 reminds you of the relationships among your design strategy, data collection activities, preliminary analysis, and hypothesis testing.

We evaluate the accuracy of hypotheses by determining the statistical likelihood that the data reveal true differences—not random sampling error. We evaluate the importance of a statistically significant difference by weighing the practical significance of any change that we measure.

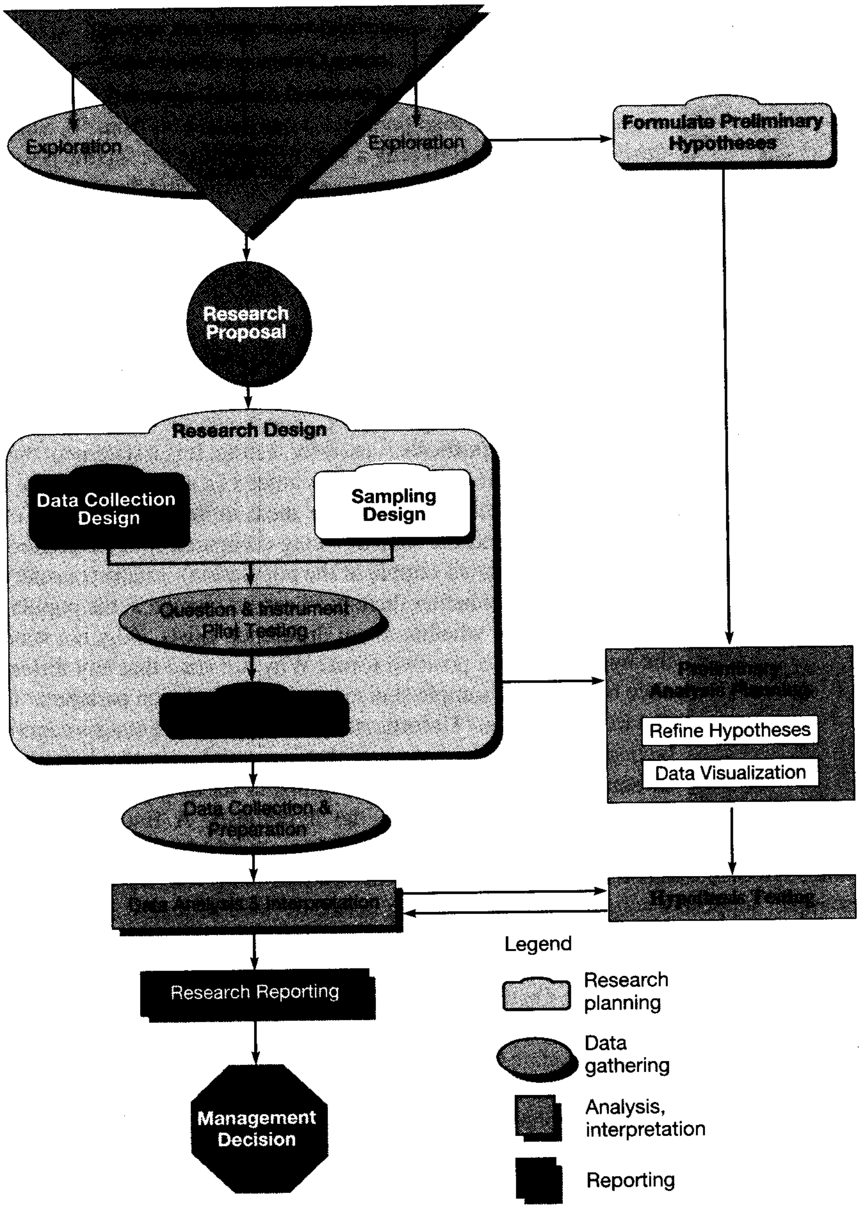
Although there are two approaches to hypothesis testing, the more established is the classical or sampling-theory approach. **Classical statistics** are found in all of the major statistics books and are widely used in research applications. This approach represents an objective view of probability in which the decision making rests totally on an analysis of available sampling data. A hypothesis is established; it is rejected or fails to be rejected, based on the sample data collected.

The second approach is known as **Bayesian statistics,** which are an extension of the classical approach. They also use sampling data, but they go beyond to consider all other available information. This additional information consists of subjective probability estimates stated in terms of degrees of belief. These subjective estimates are based on general experience rather than on specific collected data. Various decision rules are established, cost and other estimates can be introduced, and the expected outcomes of combinations of these elements are used to judge decision alternatives.

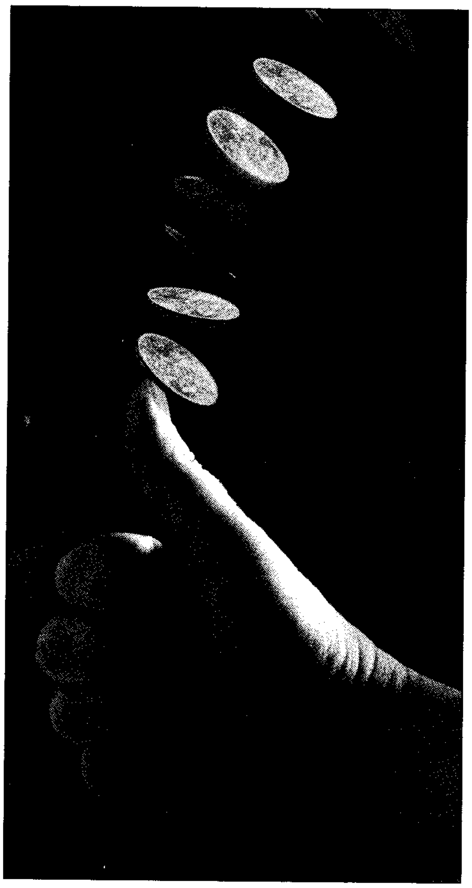
Statistical Significance

Following classical statistics approach, we accept or reject a hypothesis on the basis of sampling information alone. Since any sample will almost surely vary somewhat from its population, we must judge whether the differences are statistically significant or insignificant. A difference has **statistical significance** if there is good reason to believe the difference does not represent random sampling fluctuations only. For example, Honda, Toyota, DaimlerChrysler, Ford, and other auto companies produce hybrid vehicles using an advanced technology that combines a small gas engine with an electric motor. The vehicles run on an electric motor at

> Exhibit 18-1 Hypothesis Testing and the Research Process



slow speeds but shift to both the gasoline motor and the electric motor at city and higher freeway speeds. Their advertising strategies focus on fuel economy. Let's say that the hybrid Civic has maintained an average of about 50 miles per gallon (mpg) with a standard deviation of 10 mpg. Suppose researchers discover by analyzing all production vehicles that the mpg is now 51. Is this difference statistically significant from 50? Of course it is, because the difference is based on a *census* of the vehicles and there is no sampling involved. It has been demonstrated conclusively that the population average has moved from 50 to 51 mpg. While it is of statistical significance, whether it is of **practical significance** is another question. If a decision maker judges that this variation has no real importance, then it is of little practical significance.



Since it would be too expensive to analyze all of a manufacturer's vehicles frequently, we resort to sampling. Assume a sample of 25 cars is randomly selected and the average mpg is calculated to be 54. Is this statistically significant? The answer is not obvious. It is significant if there is good reason to believe the average mpg of the total population has moved up from 50. Since the evidence consists of only a sample, consider the second possibility: that this is only a random sampling error and thus is not significant. The task is to decide whether such a result from this sample is or is not statistically significant. To answer this question, one needs to consider further the logic of hypothesis testing.

The Logic of Hypothesis Testing

In classical tests of significance, two kinds of hypotheses are used. The **null hypothesis** is used for testing. It is a statement that no difference exists between the parameter (a measure taken by a census of the population or a prior measurement of a sample of the population) and the statistic being compared to it (a measure from a recently drawn sample of the population). Analysts usually test to determine whether there has been no change in the population of interest or whether a real difference exists. Why not state the hypothesis in a positive form? Why not state that any difference between the sample statistic and the population parameter is due to some reason? Unfortunately, this type of hypothesis cannot be tested definitively. Evidence that is consistent with a hypothesis stated in a positive form can almost never be taken as conclusive grounds for accepting the hypothesis. A finding that is consistent with this type of hypothesis might be consistent with other hypotheses too, and thus it does not demonstrate the truth of the given hypothesis.

ethical issues

For example, suppose a coin is suspected of being biased in favor of heads. The coin is flipped 100 times and the outcome is 52 heads. It would not be correct to jump to the conclusion that the coin is biased simply because more than the expected number of 50 heads resulted. The reason is that 52 heads is consistent with the hypothesis that the coin is fair. On the other hand, flipping 85 or 90 heads in 100 flips would seem to contradict the hypothesis of a fair coin. In this case there would be a strong case for a biased coin.

In the hybrid-vehicle example, the null hypothesis states that the population parameter of 50 mpg has not changed. A second, **alternative hypothesis** holds that there has been a change in average mpg (i.e., the sample statistic of 54 indicates the population value probably is no longer 50). The alternative hypothesis is the logical opposite of the null hypothesis.

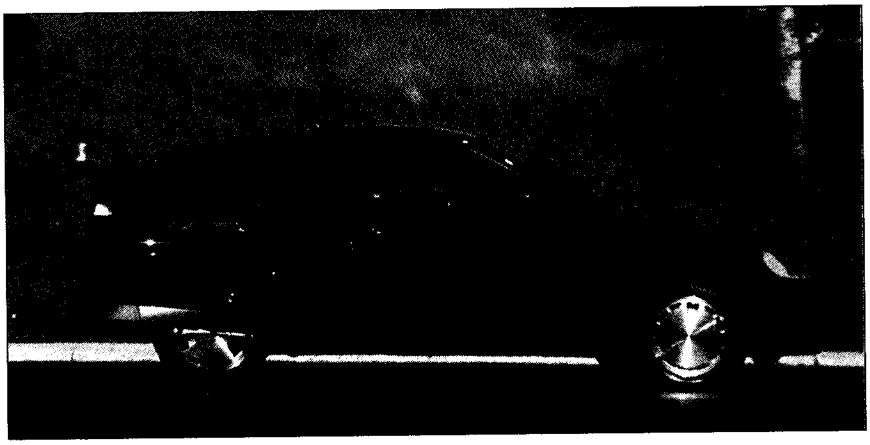
The hybrid-car example can be explored further to show how these concepts are used to test for significance:

- The null hypothesis (H_0): There has been no change from the 50 mpg average.

The alternative hypothesis (H_A) may take several forms, depending on the objective of the researchers. The H_A may be of the "not the same" or the "greater than" or "less than" form:

- The average mpg has changed from 50.
- The average mpg has increased (decreased) from 50.

These types of alternative hypotheses correspond with two-tailed and one-tailed tests. A **two-tailed test,** or *nondirectional test,* considers two possibilities: The average could be more than 50 mpg, or it could be less



Honda's Insight: a sleek, sporty, and affordable car for the environmentally conscious consumer that uses extraordinary technology to achieve remarkable gas mileage. It is powered by a lively VTEC®-E engine and an electric motor that you never have to plug in. **www.honda.com**

than 50. To test this hypothesis, the regions of rejection are divided into two tails of the distribution. A **one-tailed test,** or *directional test,* places the entire probability of an unlikely outcome into the tail specified by the alternative hypothesis. In Exhibit 18-2, the first diagram represents a nondirectional hypothesis, and the second is a directional hypothesis of the "greater than" variety.

Hypotheses for Exhibit 18-2 may be expressed in the following form:

Null $H_0: \mu = 50$ mpg

Alternative $H_A: \mu \neq 50$ mpg (not the same case)

Or

Null $H_0: \mu \leq 50$ mpg

Alternative $H_A: \mu > 50$ mpg (greater than case)

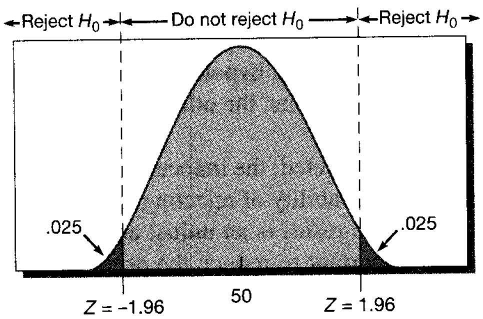
Or

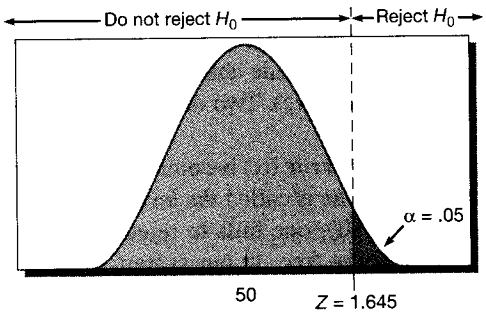
Null $H_0: \mu \geq 50$ mpg

Alternative $H_A: \mu < 50$ mpg (less than case)

In testing these hypotheses, adopt this decision rule: Take no corrective action if the analysis shows that one cannot reject the null hypothesis. Note the language "cannot reject" rather than "accept" the null hypothesis. It is argued that a null hypothesis can never be proved and therefore cannot be "accepted." Here, again, we see the influence of inductive reasoning. Unlike deduction, where the connections between premises and conclusions provide a legitimate claim of "conclusive proof," inductive conclusions do not possess that advantage. Statistical testing gives only a chance to (1) disprove (reject) or (2) fail to reject the hypothesis. Despite

> Exhibit 18-2 Two- and One-Tailed Tests at the 5% Level of Significance





>snapshot

Direct-to-Consumer Ads under Heavy Fire

Direct-to-consumer (D-to-C) pharmaceutical ads have drawn a lot of criticism since 1997 Food and Drug Administration (FDA) regulations permitted such tactics. Proponents of legislation to disallow the practice fear such ads "unfairly influence important health care decisions" by causing patients to pressure doctors and thus encourage doctors to prescribe unnecessary prescriptions. The chairman of the American Medical Association believes such advertising may create an adversarial relationship between doctor and patient. He wants to know if the ads "improve the quality of care enough to make it worth the increased costs of the medicines being advertised." One democratic legislator believes "taxpayers would not have to subsidize excessive advertising that leads to higher prices at the pharmacy counter."

Ipsos-NPD tracks this issue for the pharmaceutical industry with its monthly PharmTrends® panel, comprising 16,000 U.S. households. Panel members are measured for ad recall, prescriptions filled, physician recommendations for over-the-counter (OTC) products, and OTC products purchased, as well as condition being treated. Panel findings reveal that advertising "has encouraged higher levels of script fulfillment per year among consumers who reported that they were aware of advertising." Additionally, such advertising is credited with reminding patients to refill prescriptions. In its February InstaVue omnibus mail survey of 26,000 adults, 47 percent had seen pharmaceutical advertising in the past year, 25 percent indicated D-to-C ads encouraged them to call/visit their doctors to discuss the pharmaceutical advertised, and 15 percent reported asking for the very drug advertised.

How would you determine if this research confirmed or refuted that "pharmaceutical ads undermine quality of care"?

www.ipsos-npd.com

this terminology, it is common to hear "accept the null" rather than the clumsy "fail to reject the null." In this discussion, the less formal *accept* means "fail to reject" the null hypothesis.

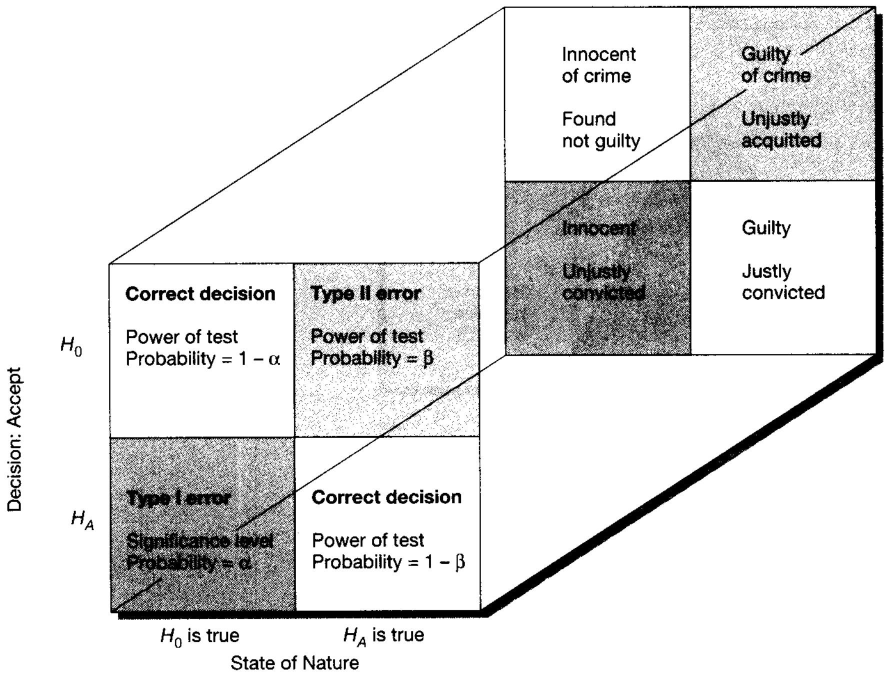
If we reject a null hypothesis (finding a statistically significant difference), then we are accepting the alternative hypothesis. In either accepting or rejecting a null hypothesis, we can make incorrect decisions. A null hypothesis can be accepted when it should have been rejected or rejected when it should have been accepted.

These problems are illustrated with an analogy to the American legal system.[1] In our system of justice, the innocence of an indicted person is presumed until proof of guilt beyond a reasonable doubt can be established. In hypothesis testing, this is the null hypothesis; there should be no difference between the presumption of innocence and the outcome unless contrary evidence is furnished. Once evidence establishes beyond reasonable doubt that innocence can no longer be maintained, a just conviction is required. This is equivalent to rejecting the null hypothesis and accepting the alternative hypothesis. Incorrect decisions or errors are the other two possible outcomes. We can unjustly convict an innocent person, or we can acquit a guilty person.

Exhibit 18-3 compares the statistical situation to the legal one. One of two conditions exists in nature—either the null hypothesis is true or the alternative hypothesis is true. An indicted person is innocent or guilty. Two decisions can be made about these conditions: One may accept the null hypothesis or reject it (thereby accepting the alternative). Two of these situations result in correct decisions; the other two lead to decision errors.

When a **Type I error** (α) is committed, a true null hypothesis is rejected; the innocent person is unjustly convicted. The value is called *the level of significance* and is the probability of rejecting the true null. With a **Type II error** (β), one fails to reject a false null hypothesis; the result is an unjust acquittal, with the guilty person going free. In our system of justice, it is more important to reduce the probability of convicting the innocent than that of acquitting the guilty. Similarly, hypothesis testing places a greater emphasis on Type I errors than on Type II errors. Next we shall examine each of these errors in more detail.

> Exhibit 18-3 Comparison of Statistical Decisions to Legal Analogy



Type I Error

Assume the hybrid manufacturer's problem is complicated by a consumer testing agency's assertion that the average mpg has changed. Assume the population mean is 50 mpg, the standard deviation of the population is 10 mpg, and the size of the sample is 25 vehicles. With this information, one can calculate the standard error of the mean ($\sigma_{\bar{X}}$) (the standard deviation of the distribution of sample means). This hypothetical distribution is pictured in Exhibit 18-4. The standard error of the mean is calculated to be 2 mpg:

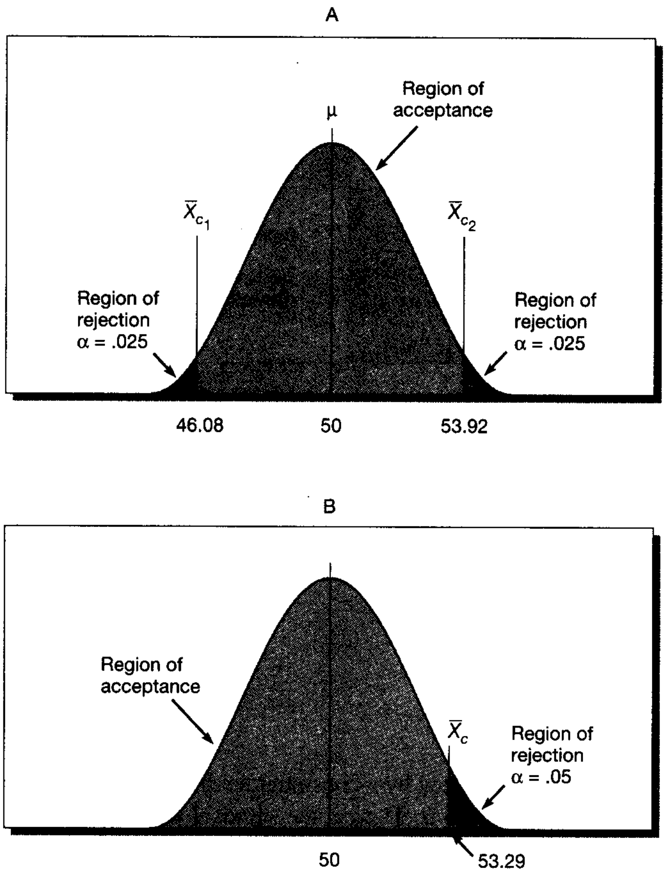
$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{25}} = 2$$

If the decision is to reject H_0 with a 95 percent confidence interval ($\alpha = .05$), a Type I error of .025 in each tail is accepted (this assumes a two-tailed test). In part A of Exhibit 18-4, see the **regions of rejection** indicated by the shaded areas. The area between these two regions is known as the **region of acceptance.** The dividing points between the rejection and acceptance areas are called **critical values.** Since the distribution of sample means is normal, the critical values can be computed in terms of the standardized random variable,[2] where

$Z = 1.96$ (significance level $= .05$)
$\bar{X}_c$ = the critical value of the sample mean
μ = the population value stated in $H_0 = 50$
$\sigma_{\bar{X}}$ = the standard error of a distribution of means of samples of 25

Thus the critical values for the test of the null hypothesis (that the mpg has not changed) are computed as follows:

> Exhibit 18-4 Probability of Making a Type I Error Given H_0 Is True



$$Z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}}$$

$$-1.96 = \frac{\bar{X}_c - 50}{2}$$

$$\bar{X}_c = 46.08$$

$$1.96 = \frac{\bar{X}_c - 50}{2}$$

$$\bar{X}_c = 53.92$$

If the probability of a Type I error is 5 percent (α = .05), the probability of a correct decision if the null hypothesis is true is 95 percent. By changing the probability of a Type I error, you move critical values either closer to or farther away from the assumed parameter of 50. This can be done if a smaller or larger α error is desired and critical values are moved to reflect this. You can also change the Type 1 error and the regions of acceptance by changing the size of the sample. For example, if you take a sample of 100, the critical values that provide a Type I error of .05 are 48.04 and 51.96.

The alternative hypothesis concerned a change in either direction from 50, but the manufacturer is interested only in increases in mpg. For this, one uses a one-tailed (greater than) H_A and places the entire region of rejection in the upper tail of the distribution. One can accept a 5 percent α risk and compute a new critical value (X_c). (See Appendix C, Exhibit C-1, to find the Z value of 1.645 for the area of .05 under the curve.) Substitute this in the Z equation and solve for $\bar{X}_c$:

$$Z = 1.645 = \frac{\bar{X}_c - 50}{2}$$

$$\bar{X}_c = 53.29$$

This new critical value, the boundary between the regions of acceptance and rejection, is pictured in part B of Exhibit 18-4.

Type II Error

The manufacturer would commit a Type II error (β) by accepting the null hypothesis ($\mu = 50$) when in truth it had changed. This kind of error is difficult to detect. The probability of committing a β error depends on five factors: (1) the true value of the parameter, (2) the α level we have selected, (3) whether a one- or two-tailed test was used to evaluate the hypothesis, (4) the sample standard deviation, and (5) the size of the sample. We secure a different β error if the new β moves from 50 to 54 rather than only to 52. We must compute separate β error estimates for each of a number of assumed new population parameters and $\bar{X}_c$ values.

To illustrate, assume μ has actually moved to 54 from 50. Under these conditions, what is the probability of our making a Type II error if the critical value is set at 53.29? (See Exhibit 18-5.) This may be expressed in the following fashion:

$$P(A_2)S_1 = \alpha = .05 \text{ (assume a one-tailed alternative hypothesis)}$$

$$P(A_1)S_2 = \beta = ?$$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{25}} = 2$$

$$Z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} = \frac{53.29 - 54}{2} = -.355$$

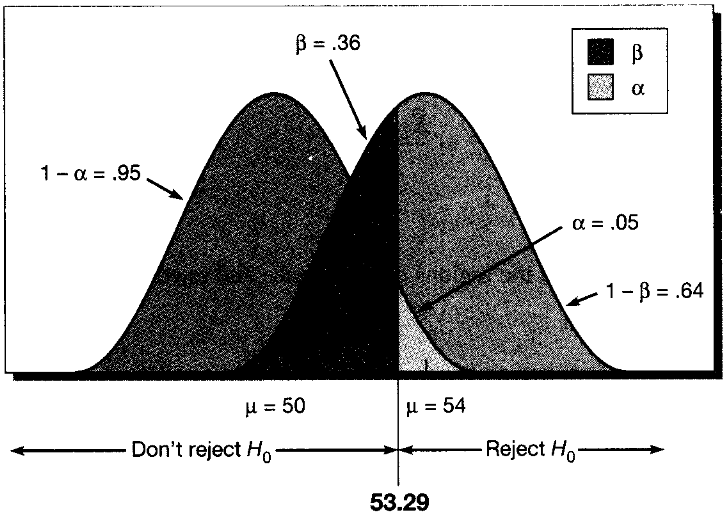
Using Exhibit C-1 in Appendix C, we interpolate between .35 and .36 Z scores to find the .355 Z score. The area between the mean and Z is .1387. β is the tail area, or the area below the Z, and is calculated as

$$\beta = .50 - .1387 = .36$$

This condition is shown in Exhibit 18-5. It is the percent of the area where we would *not* reject the null (H_0:$\mu = 50$) when, in fact, it was false because the true mean was 54. With an α of .05 and a sample of 25, there is a 36 percent probability of a Type II (β) error if the μ is 54. We also speak of the **power of the test**—that is $(1 - \beta)$. For this example, the power of the test equals 64 percent $(1 - .36)$—that is, we will correctly reject the false null hypothesis with a 64 percent probability. A power of 64 percent is less than the 80 percent minimum percentage recommended by statisticians.

There are several ways to reduce a Type II error. We can shift the critical value closer to the original μ of 50; but to do this, we must accept a bigger α. Whether to take this action depends on the evaluation of the relative α and β risks. It might be desirable to enlarge the acceptable α risk because a worsening of the mileage

> **Exhibit 18-5** Probability of Making a Type II Error



would probably call for increased efforts to stimulate efficiency. Committing a Type I error would mean only that we engaged in efforts to stimulate efficiency when the situation had not worsened. This act probably would not have many adverse effects even if mpg had not increased.

A second way to reduce Type II error is to increase sample size. For example, if the sample were increased to 100, the power of the test would be much stronger:

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{100}} = 1$$

$$Z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} = \frac{53.29 - 54}{1} = -.71$$

$$\beta = .50 - .2612 = .24$$

This would reduce the Type II error to 24 percent and increase the power of the test to 76 percent.

A third method seeks to improve both α and β errors simultaneously and is difficult to accomplish. We know that measuring instruments, observations, and recording produce error. By using a better measuring device, tightening the observation and recording processes, or devising a more efficient sample, we can reduce the variability of observations. This diminishes the standard error of estimate and in turn reduces the sampling distributions' spread. The net effect is that there is less tail area in the error regions.

Statistical Testing Procedures

Testing for statistical significance follows a relatively well-defined pattern, although authors differ in the number and sequence of steps. One six-stage sequence is as follows:

1. *State the null hypothesis.* While the researcher is usually interested in testing a hypothesis of change or differences, the null hypothesis is always used for statistical testing purposes.
2. *Choose the statistical test.* To test a hypothesis, one must choose an appropriate statistical test. There are many tests from which to choose, and there are at least four criteria that can be used in choosing a test. One is the power efficiency of the test. A more powerful test provides the same level of