

Introduction to Statistics

LEARNING OBJECTIVES

The primary objective of Chapter 1 is to introduce you to the world of statistics, thereby enabling you to:

1. List quantitative and graphical examples of statistics within a business context
2. Define important statistical terms, including population, sample, and parameter, as they relate to descriptive and inferential statistics
3. Compare the four different levels of data: nominal, ordinal, interval, and ratio

Keren Su/The Image Bank/Getty Images





Statistics Describe the State of Business in India's Countryside

India is the second largest country in the world, with more than a billion people. Nearly three-quarters of the people live in rural areas scat-

tered about the countryside in 6,000,000 villages. In fact, it may be said that 1 in every 10 people in the world live in rural India. Presently, the population in rural India can be described as poor and semi-illiterate. With an annual per capita income of less than \$1 (U.S.) per day, rural India accounts for only about one-third of total national product sales. Less than 50% of households in rural India have electricity, and many of the roads are not paved. The annual per capita consumption for toothpaste is only 30 grams per person in rural India compared to 160 grams in urban India and 400 grams in the United States.

However, in addition to the impressive size of the population, there are other compelling reasons for companies to market their goods and services to rural India. The market of rural India has been growing at five times the rate of the urban India market. There is increasing agricultural productivity, leading to growth in disposable income, and there is a reduction in the gap between the tastes of urban and rural customers. The literacy level is increasing, and people are becoming more conscious about their lifestyles and opportunities for a better life.

Nearly two-thirds of all middle-income households in India are in rural areas, with the number of middle- and high-income households in rural India expected to grow from 80 million to 111 million over the next three years. More than one-third of all rural households now have a main source of income other than farming. Virtually every home has a radio, almost 20% have a television, and more than 30% have at least one bank account.

In the early 1990s, toothpaste consumption in rural India doubled, and the consumption of shampoo increased fourfold. Recently, other products have

done well in rural India, accounting for nearly one-half of all of the country's sales of televisions, fans, bicycles, bath soap, and other products. According to MART, a New Delhi-based research organization, rural India buys 46% of all soft drinks and 49% of motorcycles sold in India. In one year alone, the market for Coca-Cola in rural India grew by 37%, accounting for 80% of new Coke drinkers in India. Because of such factors, many U.S. and Indian firms, such as Microsoft, General Electric, Kellogg's, Colgate-Palmolive, Hindustan Lever, Godrej, Nirma Chemical Works, and Mahotra Marketing, have entered the rural Indian market with enthusiasm. Marketing to rural customers often involves building categories by persuading them to try and adopt products that they may not have used before. Rural India is a huge, relatively untapped market for businesses. However, entering such a market is not without risks and obstacles. The dilemma facing companies is whether to enter this marketplace and, if so, to what extent and how.

Managerial and Statistical Questions

1. Are the statistics presented in this report exact figures or estimates?
2. How and where could the researchers have gathered such data?
3. In measuring the potential of the rural India marketplace, what other statistics could have been gathered?
4. What levels of data measurement are represented by data on rural India?
5. How can managers use these and other statistics to make better decisions about entering this marketplace?

Source: Adapted from Raja Ramachandran, "Understanding the Market Environment of India," *Business Horizons*, January 2000; P. Balakrishna and B. Sidharth, "Selling in Rural India," *The Hindu Business Line*—Internet Edition, February 16, 2004; Rohit Bansal and Srividya Easwaran, "Creative Marketing for Rural India," research paper, <http://www.indiaonline.com>; Alex Steffen, "Rural India Ain't What It Used to Be," *WorldChanging*, <http://www.worldchanging.com/archives/001235.html>; "Corporates Turn to Rural India for Growth," BS Corporate Bureau in New Delhi, August 21, 2003, <http://www.rediff.com/money/2003/aug/21rural.htm>; Rajesh Jain, "Tech Talk: The Discovery of India: Rural India," June 20, 2003, <http://www.emergic.org/archives/indi/005721.php>; "Marketing to Rural India: Making the Ends Meet," March 8, 2007, in *India Knowledge@Wharton*. <http://knowledge.wharton.upenn.edu/india/article.cfm?articleid=4172>

Every minute of the working day, decisions are made by businesses around the world that determine whether companies will be profitable and growing or whether they will stagnate and die. Most of these decisions are made with the assistance of information gathered about the marketplace, the economic and financial environment, the workforce, the competition, and other factors. Such information usually comes in the form of data or is accompanied by data.

Business statistics provides the tool through which such data are collected, analyzed, summarized, and presented to facilitate the decision-making process, and business statistics plays an important role in the ongoing saga of decision making within the dynamic world of business.



1.1 STATISTICS IN BUSINESS

Virtually every area of business uses statistics in decision making. Here are some recent examples:

- According to a TNS Retail Forward ShopperScape survey, the average amount spent by a shopper on electronics in a three-month period is \$629 at Circuit City, \$504 at Best Buy, \$246 at Wal-Mart, \$172 at Target, and \$120 at RadioShack.
- A survey of 1465 workers by Hotjobs reports that 55% of workers believe that the quality of their work is perceived the same when they work remotely as when they are physically in the office.
- A survey of 477 executives by the Association of Executive Search Consultants determined that 48% of men and 67% of women say they are more likely to negotiate for less travel compared with five years ago.
- A survey of 1007 adults by RBC Capital Markets showed that 37% of adults would be willing to drive 5 to 10 miles to save 20 cents on a gallon of gas.
- A Deloitte Retail “Green” survey of 1080 adults revealed that 54% agreed that plastic, non-compostable shopping bags should be banned.
- A recent Household Economic Survey by Statistic New Zealand determined that the average weekly household net expenditure in New Zealand was \$956 and that households in the Wellington region averaged \$120 weekly on recreation and culture. In addition, 75% of all households were satisfied or very satisfied with their material standard of living.
- The Experience’s Life After College survey of 320 recent college graduates showed that 58% moved back home after college. Thirty-two percent then remained at home for more than a year.

You can see from these few examples that there is a wide variety of uses and applications of statistics in business. Note that in most of these examples, business researchers have conducted a study and provided us rich and interesting information.

William King/The Image Bank/
Exclusive/Getty Images, Inc.



Comstock/Getty Images, Inc.



In this text we will examine several types of graphs for depicting data as we study ways to arrange or structure data into forms that are both meaningful and useful to decision makers. We will learn about techniques for sampling from a population that allow studies of the business world to be conducted more inexpensively and in a more timely manner. We will explore various ways to forecast future values and examine techniques for predicting trends. This text also includes many statistical tools for testing hypotheses and for estimating population values. These and many other exciting statistics and statistical techniques await us on this journey through business statistics. Let us begin.



BASIC STATISTICAL CONCEPTS

Business statistics, like many areas of study, has its own language. It is important to begin our study with an introduction of some basic concepts in order to understand and communicate about the subject. We begin with a discussion of the word *statistics*. The word *statistics* has many different meanings in our culture. *Webster's Third New International Dictionary* gives a comprehensive definition of **statistics** as *a science dealing with the collection, analysis, interpretation, and presentation of numerical data*. Viewed from this perspective, statistics includes all the topics presented in this text.

Photodisc/Getty Images, Inc.



The study of statistics can be organized in a variety of ways. One of the main ways is to subdivide statistics into two branches: descriptive statistics and inferential statistics. To understand the difference between descriptive and inferential statistics, definitions of *population* and *sample* are helpful. *Webster's Third New International Dictionary* defines **population** as *a collection of persons, objects, or items of interest*. The population can be a widely defined category, such as “all automobiles,” or it can be narrowly defined, such as “all Ford Mustang cars produced from 2002 to 2005.” A population can be a group of people, such as “all workers presently employed by Microsoft,” or it can be a set of objects, such as “all dishwashers produced on February 3, 2007, by the General Electric Company at the Louisville plant.” The researcher defines the population to be whatever he or she is studying. When researchers *gather data from the whole population for a given measurement of interest*, they call it a **census**. Most people are familiar with the U.S. Census. Every 10 years, the government attempts to measure all persons living in this country.

A **sample** is *a portion of the whole* and, if properly taken, is representative of the whole. For various reasons (explained in Chapter 7), researchers often prefer to work with a sample

of the population instead of the entire population. For example, in conducting quality-control experiments to determine the average life of lightbulbs, a lightbulb manufacturer might randomly sample only 75 lightbulbs during a production run. Because of time and money limitations, a human resources manager might take a random sample of 40 employees instead of using a census to measure company morale.

If a business analyst is *using data gathered on a group to describe or reach conclusions about that same group*, the statistics are called **descriptive statistics**. For example, if an instructor produces statistics to summarize a class's examination effort and uses those statistics to reach conclusions about that class only, the statistics are descriptive.

Many of the statistical data generated by businesses are descriptive. They might include number of employees on vacation during June, average salary at the Denver office, corporate sales for 2009, average managerial satisfaction score on a company-wide census of employee attitudes, and average return on investment for the Lofton Company for the years 1990 through 2008.

Another type of statistics is called **inferential statistics**. If a researcher *gathers data from a sample and uses the statistics generated to reach conclusions about the population from which the sample was taken*, the statistics are inferential statistics. The data gathered from the sample are used to infer something about a larger group. Inferential statistics are sometimes referred to as *inductive statistics*. The use and importance of inferential statistics continue to grow.

One application of inferential statistics is in pharmaceutical research. Some new drugs are expensive to produce, and therefore tests must be limited to small samples of patients. Utilizing inferential statistics, researchers can design experiments with small randomly selected samples of patients and attempt to reach conclusions and make inferences about the population.

Market researchers use inferential statistics to study the impact of advertising on various market segments. Suppose a soft drink company creates an advertisement depicting a dispensing machine that talks to the buyer, and market researchers want to measure the impact of the new advertisement on various age groups. The researcher could stratify the population into age categories ranging from young to old, randomly sample each stratum, and use inferential statistics to determine the effectiveness of the advertisement for the various age groups in the population. The advantage of using inferential statistics is that they enable the researcher to study effectively a wide range of phenomena without having to conduct a census. Most of the topics discussed in this text pertain to inferential statistics.

A *descriptive measure of the population* is called a **parameter**. Parameters are usually denoted by Greek letters. Examples of parameters are population mean (μ), population variance (σ^2), and population standard deviation (σ). A *descriptive measure of a sample* is called a **statistic**. Statistics are usually denoted by Roman letters. Examples of statistics are sample mean ($\bar{x}$), sample variance (s^2), and sample standard deviation (s).

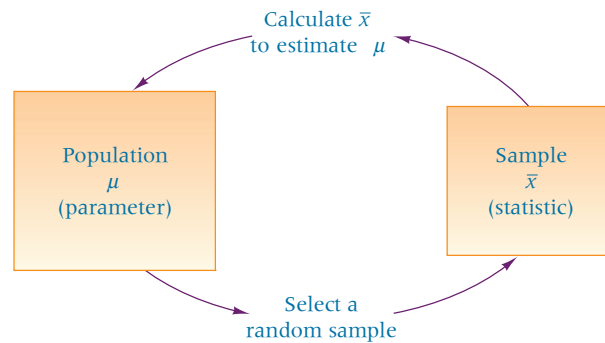
Differentiation between the terms *parameter* and *statistic* is important only in the use of inferential statistics. A business researcher often wants to estimate the value of a parameter or conduct tests about the parameter. However, the calculation of parameters is usually either impossible or infeasible because of the amount of time and money required to take a census. In such cases, the business researcher can take a random sample of the population, calculate a statistic on the sample, and infer by estimation the value of the parameter. The basis for inferential statistics, then, is the ability to make decisions about parameters without having to complete a census of the population.

For example, a manufacturer of washing machines would probably want to determine the average number of loads that a new machine can wash before it needs repairs. The parameter is the population mean or average number of washes per machine before repair. A company researcher takes a sample of machines, computes the number of washes before repair for each machine, averages the numbers, and estimates the population value or parameter by using the statistic, which in this case is the sample average. Figure 1.1 demonstrates the inferential process.

Inferences about parameters are made under uncertainty. Unless parameters are computed directly from the population, the statistician never knows with certainty whether the estimates or inferences made from samples are true. In an effort to estimate the level of confidence in the result of the process, statisticians use probability statements. For this and other reasons, part of this text is devoted to probability (Chapter 4).

FIGURE 1.1

Process of Inferential Statistics
to Estimate a Population
Mean (μ)



DATA MEASUREMENT

Millions of numerical data are gathered in businesses every day, representing myriad items. For example, numbers represent dollar costs of items produced, geographical locations of retail outlets, weights of shipments, and rankings of subordinates at yearly reviews. All such data should not be analyzed the same way statistically because the entities represented by the numbers are different. For this reason, the business researcher needs to know the *level of data measurement* represented by the numbers being analyzed.

The disparate use of numbers can be illustrated by the numbers 40 and 80, which could represent the weights of two objects being shipped, the ratings received on a consumer test by two different products, or football jersey numbers of a fullback and a wide receiver. Although 80 pounds is twice as much as 40 pounds, the wide receiver is probably not twice as big as the fullback! Averaging the two weights seems reasonable, but averaging the football jersey numbers makes no sense. The appropriateness of the data analysis depends on the level of measurement of the data gathered. The phenomenon represented by the numbers determines the level of data measurement. Four common levels of data measurement follow.

1. Nominal
2. Ordinal
3. Interval
4. Ratio

Nominal Level

The *lowest level of data measurement* is the **nominal level**. Numbers representing nominal-level data (the word *level* often is omitted) can be *used only to classify or categorize*. Employee identification numbers are an example of nominal data. The numbers are used only to differentiate employees and not to make a value statement about them. Many demographic questions in surveys result in data that are nominal because the questions are used for classification only. The following is an example of such a question that would result in nominal data:

Which of the following employment classifications best describes your area of work?

1. Educator
2. Construction worker
3. Manufacturing worker
4. Lawyer
5. Doctor
6. Other

Suppose that, for computing purposes, an educator is assigned a 1, a construction worker is assigned a 2, a manufacturing worker is assigned a 3, and so on. These numbers should be used only to classify respondents. The number 1 does not denote the top classification. It is used only to differentiate an educator (1) from a lawyer (4).

Some other types of variables that often produce nominal-level data are sex, religion, ethnicity, geographic location, and place of birth. Social Security numbers, telephone numbers, employee ID numbers, and ZIP code numbers are further examples of nominal data. Statistical techniques that are appropriate for analyzing nominal data are limited. However, some of the more widely used statistics, such as the chi-square statistic, can be applied to nominal data, often producing useful information.

Ordinal Level

Ordinal-level data measurement is higher than the nominal level. In addition to the nominal-level capabilities, ordinal-level measurement can be used to rank or order objects. For example, using ordinal data, a supervisor can evaluate three employees by ranking their productivity with the numbers 1 through 3. The supervisor could identify one employee as the most productive, one as the least productive, and one as somewhere between by using ordinal data. However, the supervisor could not use ordinal data to establish that the intervals between the employees ranked 1 and 2 and between the employees ranked 2 and 3 are equal; that is, she could not say that the differences in the amount of productivity between workers ranked 1, 2, and 3 are necessarily the same. With ordinal data, the distances or spacing represented by consecutive numbers are not always equal.

Some questionnaire Likert-type scales are considered by many researchers to be ordinal in level. The following is an example of one such scale:

This computer tutorial is	_____	_____	_____	_____	_____
	not	somewhat	moderately	very	extremely
	helpful	helpful	helpful	helpful	helpful
	1	2	3	4	5

When this survey question is coded for the computer, only the numbers 1 through 5 will remain, not the adjectives. Virtually everyone would agree that a 5 is higher than a 4 on this scale and that ranking responses is possible. However, most respondents would not consider the differences between not helpful, somewhat helpful, moderately helpful, very helpful, and extremely helpful to be equal.

Mutual funds as investments are sometimes rated in terms of risk by using measures of default risk, currency risk, and interest rate risk. These three measures are applied to investments by rating them as having high, medium, and low risk. Suppose high risk is assigned a 3, medium risk a 2, and low risk a 1. If a fund is awarded a 3 rather than a 2, it carries more risk, and so on. However, the differences in risk between categories 1, 2, and 3 are not necessarily equal. Thus, these measurements of risk are only ordinal-level measurements. Another example of the use of ordinal numbers in business is the ranking of the top 50 most admired companies in *Fortune* magazine. The numbers ranking the companies are only ordinal in measurement. Certain statistical techniques are specifically suited to ordinal data, but many other techniques are not appropriate for use on ordinal data. For example, it does not make sense to say that the average of “moderately helpful” and “very helpful” is “moderately helpful and a half.”

Because nominal and ordinal data are often derived from imprecise measurements such as demographic questions, the categorization of people or objects, or the ranking of items, *nominal and ordinal data* are **nonmetric data** and are sometimes referred to as *qualitative data*.

Interval Level

Interval-level data measurement is the *next to the highest level of data in which the distances between consecutive numbers have meaning and the data are always numerical*. The distances represented by the differences between consecutive numbers are equal; that is, interval data have equal intervals. An example of interval measurement is Fahrenheit temperature. With Fahrenheit temperature numbers, the temperatures can be ranked, and the amounts of heat between consecutive readings, such as 20°, 21°, and 22°, are the same.

In addition, with interval-level data, the zero point is a matter of convention or convenience and not a natural or fixed zero point. Zero is just another point on the scale and does not mean the absence of the phenomenon. For example, zero degrees Fahrenheit is not the lowest possible temperature. Some other examples of interval level data are the percentage change in employment, the percentage return on a stock, and the dollar change in stock price.

Ratio Level

Ratio-level data measurement is the highest level of data measurement. Ratio data have the same properties as interval data, but ratio data have an *absolute zero*, and the ratio of two numbers is meaningful. The notion of absolute zero means that zero is fixed, and the zero value in the data represents the absence of the characteristic being studied. The value of zero cannot be arbitrarily assigned because it represents a fixed point. This definition enables the statistician to create *ratios* with the data.

Examples of ratio data are height, weight, time, volume, and Kelvin temperature. With ratio data, a researcher can state that 180 pounds of weight is twice as much as 90 pounds or, in other words, make a ratio of 180:90. Many of the data gathered by machines in industry are ratio data.

Other examples in the business world that are ratio level in measurement are production cycle time, work measurement time, passenger miles, number of trucks sold, complaints per 10,000 fliers, and number of employees. With ratio-level data, no *b* factor is required in converting units from one measurement to another—that is, $y = ax$. As an example, in converting height from yards to feet: feet = 3 · yards.

Because interval- and ratio-level data are usually gathered by precise instruments often used in production and engineering processes, in national standardized testing, or in standardized accounting procedures, they are called **metric data** and are sometimes referred to as *quantitative data*.

Comparison of the Four Levels of Data

FIGURE 1.2

Usage Potential of Various Levels of Data

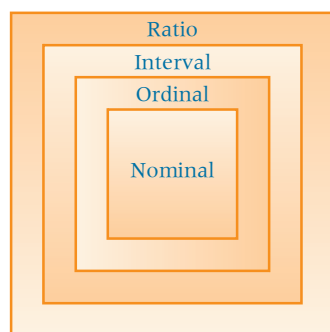


Figure 1.2 shows the relationships of the usage potential among the four levels of data measurement. The concentric squares denote that each higher level of data can be analyzed by any of the techniques used on lower levels of data but, in addition, can be used in other statistical techniques. Therefore, ratio data can be analyzed by any statistical technique applicable to the other three levels of data plus some others.

Nominal data are the most limited data in terms of the types of statistical analysis that can be used with them. Ordinal data allow the researcher to perform any analysis that can be done with nominal data and some additional analyses. With ratio data, a statistician can make ratio comparisons and appropriately do any analysis that can be performed on nominal, ordinal, or interval data. Some statistical techniques require ratio data and cannot be used to analyze other levels of data.

Statistical techniques can be separated into two categories: parametric statistics and nonparametric statistics. **Parametric statistics** require that data be interval or ratio. If the data are nominal or ordinal, **nonparametric statistics** must be used. Nonparametric statistics can also be used to analyze interval or ratio data. This text focuses largely on parametric statistics, with the exception of Chapter 16 and Chapter 17, which contain nonparametric techniques. Thus much of the material in this text requires that data be interval or ratio data.

DEMONSTRATION PROBLEM 1.1



Demonstration Problem

Many changes continue to occur in the healthcare industry. Because of increased competition for patients among providers and the need to determine how providers can better serve their clientele, hospital administrators sometimes administer a quality satisfaction survey to their patients after the patient is released. The following types of

questions are sometimes asked on such a survey. These questions will result in what level of data measurement?

1. How long ago were you released from the hospital?
2. Which type of unit were you in for most of your stay?
 - ☐ Coronary care
 - ☐ Intensive care
 - ☐ Maternity care
 - ☐ Medical unit
 - ☐ Pediatric/children's unit
 - ☐ Surgical unit
3. In choosing a hospital, how important was the hospital's location?

(circle one)

Very Important	Somewhat Important	Not Very Important	Not at All Important
-------------------	-----------------------	-----------------------	-------------------------
4. How serious was your condition when you were first admitted to the hospital?

☐ Critical	☐ Serious	☐ Moderate	☐ Minor
-----------------------------------	----------------------------------	-----------------------------------	--------------------------------
5. Rate the skill of your doctor:

☐ Excellent	☐ Very Good	☐ Good	☐ Fair	☐ Poor
------------------------------------	------------------------------------	-------------------------------	-------------------------------	-------------------------------

Solution

Question 1 is a time measurement with an absolute zero and is therefore ratio-level measurement. A person who has been out of the hospital for two weeks has been out twice as long as someone who has been out of the hospital for one week.

Question 2 yields nominal data because the patient is asked only to categorize the type of unit he or she was in. This question does not require a hierarchy or ranking of the type of unit. Questions 3, 4, and 5 are likely to result in ordinal-level data. Suppose a number is assigned the descriptors in each of these three questions. For question 3, "very important" might be assigned a 4, "somewhat important" a 3, "not very important" a 2, and "not at all important" a 1. Certainly, the higher the number, the more important is the hospital's location. Thus, these responses can be ranked by selection. However, the increases in importance from 1 to 2 to 3 to 4 are not necessarily equal. This same logic applies to the numeric values assigned in questions 4 and 5.

STATISTICS IN BUSINESS TODAY

Cellular Phone Use in Japan

The Communications and Information Network Association of Japan (CIAJ) conducts an annual study of cellular phone use in Japan. A recent survey was taken as part of this study using a sample of 600 cell phone users split evenly between men and women and almost equally distributed over six age brackets. The survey was administered in the greater Tokyo and Osaka metropolitan areas. The study produced several interesting findings. It was determined that 62.2% had replaced their handsets in the previous 10 months. A little more than 6% owned a second cell phone. Of these, the objective of about two-thirds was to own one for business use and a second one

for personal use. Of all those surveyed, 18.2% used their handsets to view videos, and another 17.3% were not currently using their handsets to view videos but were interested in doing so. Some of the everyday uses of cell phones included e-mailing (91.7% of respondents), camera functions (77.7%), Internet searching (46.7%), and watching TV (28.0%). In the future, respondents hoped there would be cell phones with high-speed data transmission that could be used to send and receive PC files (47.7%), for video services such as You Tube (46.9%), for downloading music albums (45.3%) and music videos (40.8%), and for downloading long videos such as movies (39.2%).

Statistical Analysis Using the Computer: Excel and Minitab

The advent of the modern computer opened many new opportunities for statistical analysis. The computer allows for storage, retrieval, and transfer of large data sets. Furthermore, computer software has been developed to analyze data by means of sophisticated statistical techniques. Some widely used statistical techniques, such as multiple regression, are so tedious and cumbersome to compute manually that they were of little practical use to researchers before computers were developed.

Business statisticians use many popular statistical software packages, including Minitab, SAS, and SPSS. Many computer spreadsheet software packages also have the capability of analyzing data statistically. In this text, the computer statistical output presented is from both the Minitab and the Microsoft Excel software.



Statistics Describe the State of Business in India's Countryside



Several statistics were reported in the Decision Dilemma about rural India, including the average annual consumption of toothpaste per person, the percentage of households having electricity, and the percentage of households that have at least one bank account. The authors of the sources from which the Decision Dilemma was drawn never stated whether the reported statistics were based on actual data drawn from a census of rural India households or were based on estimates taken from a sample of rural households. If the data came from a census, then the totals, averages, and percentages presented in the Decision Dilemma are parameters. If, on the other hand, the data were gathered from samples, then they are statistics. Although governments especially do conduct censuses and at least some of the reported numbers could be parameters, more often than not, such data are gathered from samples of people or items. For example, in rural India, the government, academicians, or business researchers could have taken random samples of households, gathering consumer statistics that are then used to estimate population parameters, such as percentage of households with televisions, and so forth.

In conducting research on a topic like consumer consumption in rural India, there is potential for a wide variety of statistics to be gathered that represent several levels of data. For example, ratio-level measurements on items such

as income, number of children, age of household heads, number of livestock, and grams of toothpaste consumed per year might be obtained. On the other hand, if researchers use a Likert scale (1-to-5 measurements) to gather responses about the interests, likes, and preferences of rural India consumers, an ordinal-level measurement would be obtained, as would the ranking of products or brands in market research studies. Other variables, such as geographic location, sex, occupation, or religion, are usually measured with nominal data.

The decision to enter the rural India market is not just a marketing decision. It involves production capacity and schedule issues, transportation challenges, financial commitments, managerial growth or reassignment, accounting issues (accounting for rural India may differ from techniques used in traditional markets), information systems, and other related areas. With so much on the line, company decision makers need as much relevant information available as possible. In this Decision Dilemma, it is obvious to the decision maker that rural India is still quite poor and illiterate. Its capacity as a market is great. The statistics on the increasing sales of a few personal-care products look promising. What are the future forecasts for the earning power of people in rural India? Will major cultural issues block the adoption of the types of products that companies want to sell there? The answers to these and many other interesting and useful questions can be obtained by the appropriate use of statistics. The 750 million people living in rural India represent the second largest group of people in the world. It certainly is a market segment worth studying further.

ETHICAL CONSIDERATIONS

With the abundance and proliferation of statistical data, potential misuse of statistics in business dealings is a concern. It is, in effect, unethical business behavior to use statistics out of context. Unethical businesspeople might use only selective data from studies to underscore their point, omitting statistics from the same studies that argue against their case. The results of statistical studies can be misstated or overstated to gain favor.

This chapter noted that if data are nominal or ordinal, then only nonparametric statistics are appropriate for

analysis. The use of parametric statistics to analyze nominal and/or ordinal data is wrong and could be considered under some circumstances to be unethical.

In this text, each chapter contains a section on ethics that discusses how businesses can misuse the techniques presented in the chapter in an unethical manner. As both users and producers, business students need to be aware of the potential ethical pitfalls that can occur with statistics.

SUMMARY

Statistics is an important decision-making tool in business and is used in virtually every area of business. In this course, the word *statistics* is defined as the science of gathering, analyzing, interpreting, and presenting data.

The study of statistics can be subdivided into two main areas: *descriptive statistics* and *inferential statistics*. Descriptive statistics result from gathering data from a body, group, or population and reaching conclusions only about that group. Inferential statistics are generated from the process of gathering sample data from a group, body, or population and reaching conclusions about the larger group from which the sample was drawn.

The appropriate type of statistical analysis depends on the level of data measurement, which can be (1) *nominal*, (2) *ordinal*, (3) *interval*, or (4) *ratio*. Nominal is the lowest level, representing classification of only such data as geographic location, sex, or Social Security number. The next level is ordi-

nal, which provides rank ordering measurements in which the intervals between consecutive numbers do not necessarily represent equal distances. Interval is the next to highest level of data measurement in which the distances represented by consecutive numbers are equal. The highest level of data measurement is ratio, which has all the qualities of interval measurement, but ratio data contain an absolute zero and ratios between numbers are meaningful. Interval and ratio data sometimes are called *metric* or *quantitative* data. Nominal and ordinal data sometimes are called *nonmetric* or *qualitative* data.

Two major types of inferential statistics are (1) *parametric statistics* and (2) *nonparametric statistics*. Use of parametric statistics requires interval or ratio data and certain assumptions about the distribution of the data. The techniques presented in this text are largely parametric. If data are only nominal or ordinal in level, nonparametric statistics must be used.

KEY TERMS



Flash Cards

census

descriptive statistics

inferential statistics
interval-level data
metric data
nominal-level data
nonmetric data
nonparametric statistics

ordinal-level data	statistic
parameter	statistics
parametric statistics	
population	
ratio-level data	
sample	

SUPPLEMENTARY PROBLEMS

- 1.1

Give a specific example of data that might be gathered from each of the following business disciplines: accounting, finance, human resources, marketing, information systems, production, and management. An example in the marketing area might be “number of sales per month by each salesperson.”
- 1.2

State examples of data that can be gathered for decision making purposes from each of the following industries: manufacturing, insurance, travel, retailing, communications, computing, agriculture, banking, and healthcare. An example in the travel industry might be the cost of business travel per day in various European cities.

- 1.3 Give an example of *descriptive* statistics in the recorded music industry. Give an example of how *inferential* statistics could be used in the recorded music industry. Compare the two examples. What makes them different?
- 1.4 Suppose you are an operations manager for a plant that manufactures batteries. Give an example of how you could use *descriptive* statistics to make better managerial decisions. Give an example of how you could use *inferential* statistics to make better managerial decisions.
- 1.5 Classify each of the following as nominal, ordinal, interval, or ratio data.
- The time required to produce each tire on an assembly line
 - The number of quarts of milk a family drinks in a month
 - The ranking of four machines in your plant after they have been designated as excellent, good, satisfactory, and poor
 - The telephone area code of clients in the United States
 - The age of each of your employees
 - The dollar sales at the local pizza shop each month
 - An employee's identification number
 - The response time of an emergency unit
- 1.6 Classify each of the following as nominal, ordinal, interval, or ratio data.
- The ranking of a company by *Fortune* 500
 - The number of tickets sold at a movie theater on any given night
 - The identification number on a questionnaire
 - Per capita income
 - The trade balance in dollars
 - Profit/loss in dollars
 - A company's tax identification
 - The Standard & Poor's bond ratings of cities based on the following scales:

Rating	Grade
Highest quality	AAA
High quality	AA
Upper medium quality	A
Medium quality	BBB
Somewhat speculative	BB
Low quality, speculative	B
Low grade, default possible	CCC
Low grade, partial recovery possible	CC
Default, recovery unlikely	C

- 1.7 The Rathburn Manufacturing Company makes electric wiring, which it sells to contractors in the construction industry. Approximately 900 electric contractors purchase wire from Rathburn annually. Rathburn's director of marketing wants to determine electric contractors' satisfaction with Rathburn's wire. He developed a questionnaire that yields a satisfaction score between 10 and 50 for participant responses. A random sample of 35 of the 900 contractors is asked to complete a satisfaction survey. The satisfaction scores for the 35 participants are averaged to produce a mean satisfaction score.
- What is the population for this study?
 - What is the sample for this study?
 - What is the statistic for this study?
 - What would be a parameter for this study?

ANALYZING THE DATABASES

see www.wiley.com/college/black



Nine databases are available with this text, providing additional opportunities to apply the statistics

presented in this course. These databases are located in WileyPLUS, and each is available in either Minitab or Excel format for your convenience. These nine databases represent a wide variety of business areas, such as agribusiness, consumer spending, energy, finance, healthcare, international labor, manufacturing, and the stock market. Altogether, these databases contain 61 variables and 7722 observations. The data are gathered from such reliable sources as the U.S. government's Bureau of Labor, the U.S. Department of Agriculture, the American Hospital Association, the Energy Information Administration, *Moody's Handbook of Common Stocks*, and the U.S. Census Bureau. Five of the nine databases contain time-series data. These databases are:

12-YEAR GASOLINE DATABASE

The 12-year time-series gasoline database contains monthly data for four variables: U.S. Gasoline Prices, OPEC Spot Price,

U.S. Finished Motor Gasoline Production, and U.S. Natural Gas Wellhead Price. There are 137 data entries for each variable. U.S. Gasoline Prices are given in cents, the OPEC Spot Price is given in dollars per barrel, U.S. Finished Motor Gasoline Production is given in 1000 barrels per day, and U.S. Natural Gas Wellhead Price is given in dollars per 1000 cubic feet.

CONSUMER FOOD DATABASE

The consumer food database contains five variables: Annual Food Spending per Household, Annual Household Income, Non-Mortgage Household Debt, Geographic Region of the U.S. of the Household, and Household Location. There are 200 entries for each variable in this database representing 200 different households from various regions and locations in the United States. Annual Food Spending per Household, Annual Household Income, and Non-Mortgage Household Debt are all given in dollars. The variable, Region, tells in which one of four regions the household resides. In this variable, the Northeast is coded as 1, the Midwest is coded 2, the South is coded as 3, and the West is coded as 4. The variable, Location, is coded as 1 if the household is in a metropolitan

area and 2 if the household is outside a metro area. The data in this database were randomly derived and developed based on actual national norms.

MANUFACTURING DATABASE

This database contains eight variables taken from 20 industries and 140 subindustries in the United States. Some of the industries are food products, textile mill products, furniture, chemicals, rubber products, primary metals, industrial machinery, and transportation equipment. The eight variables are Number of Employees, Number of Production Workers, Value Added by Manufacture, Cost of Materials, Value of Industry Shipments, New Capital Expenditures, End-of-Year Inventories, and Industry Group. Two variables, Number of Employees and Number of Production Workers, are in units of 1,000. Four variables, Value Added by Manufacture, Cost of Materials, New Capital Expenditures, and End-of-Year Inventories, are in million-dollar units. The Industry Group variable consists of numbers from 1 to 20 to denote the industry group to which the particular subindustry belongs. Value of Industry Shipments has been recoded to the following 1-to-4 scale.

- 1 = \$0 to \$4.9 billion
- 2 = \$5 billion to \$13.9 billion
- 3 = \$14 billion to \$28.9 billion
- 4 = \$29 billion or more

INTERNATIONAL LABOR DATABASE

This time-series database contains the civilian unemployment rates in percent from seven countries presented yearly over a 49-year period. The data are published by the Bureau of Labor Statistics of the U.S. Department of Labor. The countries are the United States, Canada, Australia, Japan, France, Germany, and Italy.

FINANCIAL DATABASE

The financial database contains observations on eight variables for 100 companies. The variables are Type of Industry, Total Revenues (\$ millions), Total Assets (\$ millions), Return on Equity (%), Earnings per Share (\$), Average Yield (%), Dividends per Share (\$), and Average Price per Earnings (P/E) ratio. The companies represent seven different types of industries. The variable “Type” displays a company’s industry type as:

- 1 = apparel
- 2 = chemical
- 3 = electric power
- 4 = grocery
- 5 = healthcare products
- 6 = insurance
- 7 = petroleum

ENERGY DATABASE

The time-series energy database consists of data on five energy variables over a period of 35 years. The five variables are U.S. Energy Consumption, World Crude Oil Production, U.S. Nuclear Electricity Generation, U.S. Coal Production, and U.S. Natural Dry Gas Production. U.S. Energy Consumption

is given in quadrillion BTUs per year, World Crude Oil Production is given in million barrels per day, U.S. Nuclear Electricity Generation is given in billion kilowatt-hours, U.S. Coal Production is given in million short tons, and U.S. Natural Dry Gas Production is given in million cubic feet.

U.S. AND INTERNATIONAL STOCK MARKET DATABASE

This database contains seven variables—three from the U.S. stock market and four from international stock markets—with data representing monthly averages of each over a period of five years resulting in 60 data points per variable. The U.S. stock market variables include the Dow Jones Industrial Average, the NASDAQ, and Standard & Poor’s 500. The four international stock market variables of Nikkei 225, Hang Seng, FTSE 100, and IPC represent Japan, Hong Kong, United Kingdom, and Mexico.

HOSPITAL DATABASE

This database contains observations for 11 variables on U.S. hospitals. These variables include Geographic Region, Control, Service, Number of Beds, Number of Admissions, Census, Number of Outpatients, Number of Births, Total Expenditures, Payroll Expenditures, and Personnel.

The region variable is coded from 1 to 7, and the numbers represent the following regions:

- 1 = South
- 2 = Northeast
- 3 = Midwest
- 4 = Southwest
- 5 = Rocky Mountain
- 6 = California
- 7 = Northwest

Control is a type of ownership. Four categories of control are included in the database:

- 1 = government, nonfederal
- 2 = nongovernment, not-for-profit
- 3 = for-profit
- 4 = federal government

Service is the type of hospital. The two types of hospitals used in this database are:

- 1 = general medical
- 2 = psychiatric

The total expenditures and payroll variables are in units of \$1,000.

AGRIBUSINESS TIME-SERIES DATABASE

The agribusiness time-series database contains the monthly weight (in 1,000 lbs.) of cold storage holdings for six different vegetables and for total frozen vegetables over a 14-year period. Each of the seven variables represents 168 months of data. The six vegetables are green beans, broccoli, carrots, sweet corn, onions, and green peas. The data are published by the National Agricultural Statistics Service of the U.S. Department of Agriculture.

ASSIGNMENT

Use the databases to answer the following questions.

1. In the manufacturing database, what is the level of data for each of the following variables?
 - a. Number of Production Workers
 - b. Cost of Materials
 - c. Value of Industry Shipments
 - d. Industry Group
2. In the hospital database, what is the level of data for each of the following variables?
 - a. Region
 - b. Control
 - c. Number of Beds
 - d. Personnel
3. In the financial database, what is the level of data for each of the following variables?
 - a. Type of Industry
 - b. Total Assets
 - c. P/E Ratio

CASE**DIGIORNO PIZZA: INTRODUCING A FROZEN PIZZA TO COMPETE WITH CARRY-OUT**

Kraft Foods successfully introduced DiGiorno Pizza into the marketplace in 1996, with first year sales of \$120 million, followed by \$200 million in sales in 1997. It was neither luck nor coincidence that DiGiorno Pizza was an instant success. Kraft conducted extensive research about the product and the marketplace before introducing this product to the public. Many questions had to be answered before Kraft began production. For example, why do people eat pizza? When do they eat pizza? Do consumers believe that carry out pizza is always more tasty?

SMI-Alcott conducted a research study for Kraft in which they sent out 1,000 surveys to pizza lovers. The results indicated that people ate pizza during fun social occasions or at home when no one wanted to cook. People used frozen pizza mostly for convenience but selected carry out pizza for a variety of other reasons, including quality and the avoidance of cooking. The Loran Marketing Group conducted focus groups for Kraft with women aged 25 to 54. Their findings showed that consumers used frozen pizza for convenience but wanted carry out pizza taste. Kraft researchers realized that if they were to launch a successful frozen pizza that could compete with carry out pizza, they had to develop a frozen pizza that (a) had restaurant take-out quality, (b) possessed flavor variety, (c) was fast and easy to prepare, and (d) had the convenience of freezer storage. To satisfy these seemingly divergent goals, Kraft developed DiGiorno Pizza, which rises in the oven as it cooks. This impressed focus group members, and in a series of blind taste tests conducted by Product Dynamics, DiGiorno Pizza beat out all frozen pizzas and finished second overall behind one carry out brand.

DiGiorno Pizza has continued to grow in sales and market share over the years. By 2005, sales had topped the \$600 million mark, and DiGiorno Pizza held nearly a quarter of the market share of frozen pizza sales. In 2004, Kraft successfully introduced DiGiorno thin, crispy-crust pizza into the market. Sales revenues continue to grow for this leading brand of frozen pizza.

Discussion

Think about the market research that was conducted by Kraft and the fact that they used several companies.

1. What are some of the populations that Kraft might have been interested in measuring for these studies? Did Kraft actually attempt to contact entire populations? What samples were taken? In light of these two questions, how was the inferential process used by Kraft in their market research? Can you think of any descriptive statistics that might have been used by Kraft in their decision-making process?
2. In the various market research efforts made by Kraft for DiGiorno, some of the possible measurements appear in the following list. Categorize these by level of data. Think of some other measurements that Kraft researchers might have made to help them in this research effort, and categorize them by level of data.
 - a. Number of pizzas consumed per week per household
 - b. Age of pizza purchaser
 - c. Zip code of the survey respondent
 - d. Dollars spent per month on pizza per person
 - e. Time in between purchases of pizza
 - f. Rating of taste of a given pizza brand on a scale from 1 to 10, where 1 is very poor tasting and 10 is excellent taste
 - g. Ranking of the taste of four pizza brands on a taste test
 - h. Number representing the geographic location of the survey respondent
 - i. Quality rating of a pizza brand as excellent, good, average, below average, poor
 - j. Number representing the pizza brand being evaluated
 - k. Sex of survey respondent

Source: Adapted from "Upper Crust," *American Demographics*, March 1999, p. 58; *Marketwatch—News That Matters* Web sites, "What's in a Name? Brand Extension Potential" and "DiGiorno Rising Crust Delivers \$200 Million," formerly at <http://www.foodexplorer.com/BUSINESS/Products/MarketAnalysis/PF02896b.htm>, last accessed in 1999. Web site for Kraft's 2005 Investor Meeting: Changing Consumers, Changing Kraft, May 25, http://media.corporateir.net/media_files/nys/kft/presentations/kft_050510c.pdf.