INTRODUCTION

Suppose you work for a large company and your supervisor asks you to decide if a new version of a smartphone should be produced and sold. You start by thinking about the product's innovations and new features. Then, you stop and realize the weight of the decision. The product will need to make a profit so the pricing and the costs of production and distribution are all very important. The decision to introduce the product is based on many alternatives. So how will you know? Where do you start?

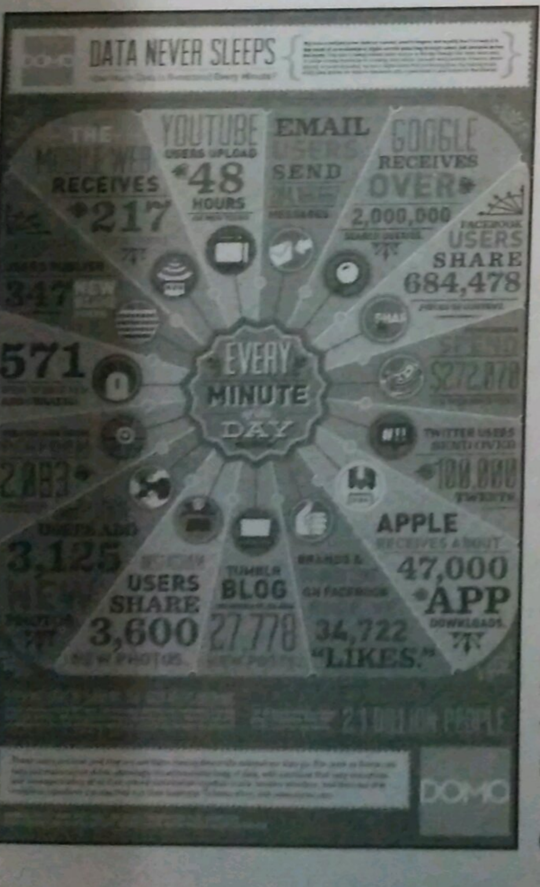
Without a long experience in the industry, beginning to develop an intelligence that will make you an expert is essential. You select three other people to work with and meet with them. The conversation focuses on what you need to know and what information and data you need. In your meeting, many questions are asked. How many competitors are already in the market? How are smartphones priced? What design features do competitors' products have? What features does the market require? What do customers want in a smartphone? What do customers like about the existing products? The answers will be based on business intelligence consisting of data and information collected through customer surveys, engineering analysis, and market research. In the end, your presentation to support your decision regarding the introduction of a new smartphone is based on the statistics that you use to summarize and organize your data, the statistics that you use to compare the new product to existing products, and the statistics to estimate future sales, costs, and revenues. The statistics will be the focus of the conversation that you will have with your supervisor about this very important decision.

As a decision maker, you will need to acquire and analyze data to support your decisions. The purpose of this text is to develop your knowledge of basic statistical techniques and methods and how to apply them to develop the business and personal intelligence that will help you make decisions.

LO1-1
Explain why knowledge of statistics is important.

WHY STUDY STATISTICS?

If you look through your university catalogue, you will find that statistics is required for many college programs. As you investigate a future career in accounting, economics, human resources, finance, or other business area, you will also discover that statistics is required as part of these college programs. So why is an education in statistics a requirement in so many disciplines?



A major driver of the requirement for statistics knowledge is the technologies available for capturing data. Examples include the technology that Google uses to track how Internet users access websites. As people use Google to search the Internet, Google records every search and then uses these data to sort and prioritize the results for future Internet searches. One recent estimate indicates that Google processes 20,000 terabytes of information per day. Big-box retailers like Target, Walmart, Kroger, and others scan every purchase and use the data to manage the distribution of products, to make decisions about marketing and sales, and to track daily and even hourly sales. Police departments collect and use data to provide city residents with maps that communicate information about crimes committed and their location. Every organization is collecting and using data to develop knowledge and intelligence that will help people make informed decisions, and to track the implementation of their decisions. The graphic to the left shows the amount of data generated every minute (www.domo.com). A good working knowledge of statistics is useful for summarizing and organizing data to provide information that is useful and supportive of decision making. Statistics is used to make valid comparisons and to predict the outcomes of decisions.

In summary, there are at least three reasons for studying statistics: (1) data are collected everywhere and require statistical knowledge to make the information useful, (2) statistical techniques are used to make professional and personal decisions, and (3) no matter what your career, you will need a knowledge of statistics to understand the world and to be conversant in your career. An understanding of statistics and statistical method will help you make more effective personal and professional decisions.

LO1-2
Define statistics and provide an example of how statistics is applied.

WHAT IS MEANT BY STATISTICS?

This question can be rephrased in two, subtly different ways: what are statistics and what is statistics? To answer the first question, a statistic is a number used to communicate a piece of information. Examples of **statistics** are:

- The inflation rate is 2%.
- Your grade point average is 3.5.
- The price of a new Tesla premium electric sedan is $85,400.

Each of these statistics is a numerical fact and communicates a very limited piece of information that is not very useful by itself. However, if we recognize that each of these statistics is part of a larger discussion, then the question "what **is** statistics" is applicable. Statistics is the set of knowledge and skills used to organize, summarize, and analyze data. The results of statistical analysis will start interesting conversations in the search for knowledge and intelligence that will help us make decisions. For example:

- The inflation rate for the calendar year was 2%. By applying statistics we could compare this year's inflation rate to the past observations of inflation. Is it higher, lower, or about the same? Is there a trend of increasing or decreasing inflation? Is there a relationship between interest rates and government bonds?
- Your grade point average (GPA) is 3.5. By collecting data and applying statistics, you can determine the required GPA to be admitted to the Master of Business Administration program at the University of Chicago, Harvard, or the University of Michigan. You can determine the likelihood that you would be admitted to a particular program. You may be interested in interviewing for a management position with Procter & Gamble. What GPA does Procter & Gamble require for college graduates with a bachelor's degree? Is there a range of acceptable GPAs?
- You are budgeting for a new car. You would like to own an electric car with a small carbon footprint. The price for the Tesla premium electric sedan is $85,400. By collecting additional data and applying statistics, you can analyze the alternatives. For example, another choice is a hybrid car that runs on both gas and electricity such as a Toyota Prius. It can be purchased for about $27,000. Another hybrid, the Chevrolet Volt, costs about $32,000. What are the differences in the cars' specifications? What additional information can be collected and summarized so that you can make a good purchase decision?

Another example of using statistics to provide information to evaluate decisions is the distribution and market share of Frito-Lay products. Data are collected on each of the Frito-Lay product lines. These data include the market share and the pounds of product sold. Statistics is used to present this information in a bar chart in Chart 1–1. It clearly shows Frito-Lay's dominance in the potato, corn, and tortilla chip markets. It also shows the absolute measure of pounds of each product line consumed in the United States.

These examples show that statistics is more than the presentation of numerical information. Statistics is about collecting and processing information to create a conversation, to stimulate additional questions, and to provide a basis for making decisions. Specifically, we define statistics as:

> **STATISTICS** The science of collecting, organizing, presenting, analyzing, and interpreting data to assist in making more effective decisions.

STATISTICS IN ACTION

We call your attention to a feature of our textbook—*Statistics in Action*. Read each one carefully to get an appreciation of the wide application of statistics in management, economics, nursing, law enforcement, sports, and other disciplines.

- In 2013, *Forbes* published a list of the richest Americans. William Gates, founder of Microsoft Corporation, is the richest. His net worth is estimated at $66.0 billion. (www.forbes.com)
- In 2013, the four largest privately owned American companies, ranked by revenue, were Cargill, Koch Industries, Mars, and Bechtel. (www.forbes.com)
- In the United States, a typical high school graduate earns $652 per week, a typical college graduate with a bachelor's degree earns $1,066 per week, and a typical college graduate with a master's degree earns $1,300 per week. (www.bls.gov/emp/ep_chart_001.htm)

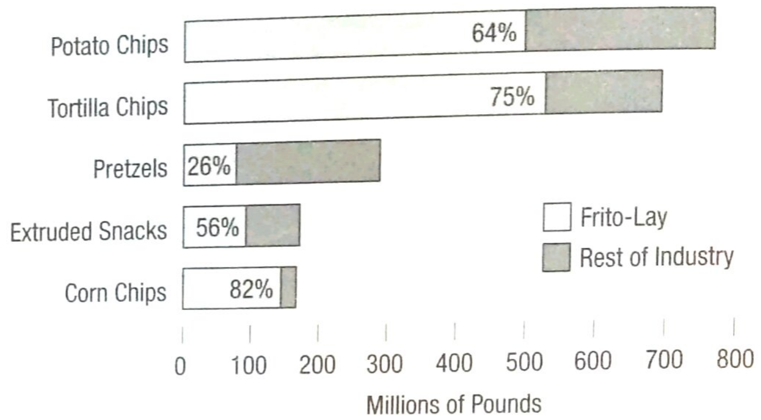


CHART 1–1 Frito-Lay Volume and Share of Major Snack Chip Categories in U.S. Supermarkets

In this book, you will learn the basic techniques and applications of statistics that you can use to support your decisions, both personal and professional. To start, we will differentiate between descriptive and inferential statistics.

LO1-3
Differentiate between descriptive and inferential statistics.

TYPES OF STATISTICS

When we use statistics to generate information for decision making from data, we use either descriptive statistics or inferential statistics. Their application depends on the questions asked and the type of data available.

Descriptive Statistics

Masses of unorganized data—such as the census of population, the weekly earnings of thousands of computer programmers, and the individual responses of 2,000 registered voters regarding their choice for president of the United States—are of little value as is. However, descriptive statistics can be used to organize data into a meaningful form. We define **descriptive statistics** as:

> **DESCRIPTIVE STATISTICS** Methods of organizing, summarizing, and presenting data in an informative way.

The following are examples that apply descriptive statistics to summarize a large amount of data and provide information that is easy to understand.

- There are a total of 46,837 miles of interstate highways in the United States. The interstate system represents only 1% of the nation's total roads but carries more than 20% of the traffic. The longest is I-90, which stretches from Boston to Seattle, a distance of 3,099 miles. The shortest is I-878 in New York City, which is 0.70 mile in length. Alaska does not have any interstate highways, Texas has the most interstate miles at 3,232, and New York has the most interstate routes with 28.
- The average person spent $103.00 on traditional Valentine's Day merchandise in 2013. This is an increase of $0.50 from 2012. As in previous years, men spent nearly twice the amount women spent on the holiday. The average man spent $135.35 to impress the people in his life while women only spent $72.28. Family pets also feel the love; the average person spent $3.27 on his or her furry friends, up from $2.17 last year.

Statistical methods and techniques to generate descriptive statistics are presented in Chapters 2 and 4. These include organizing and summarizing data with frequency distributions and presenting frequency distributions with charts and graphs.

In addition, statistical measures to summarize the characteristics of a distribution are discussed in Chapter 3.

Inferential Statistics

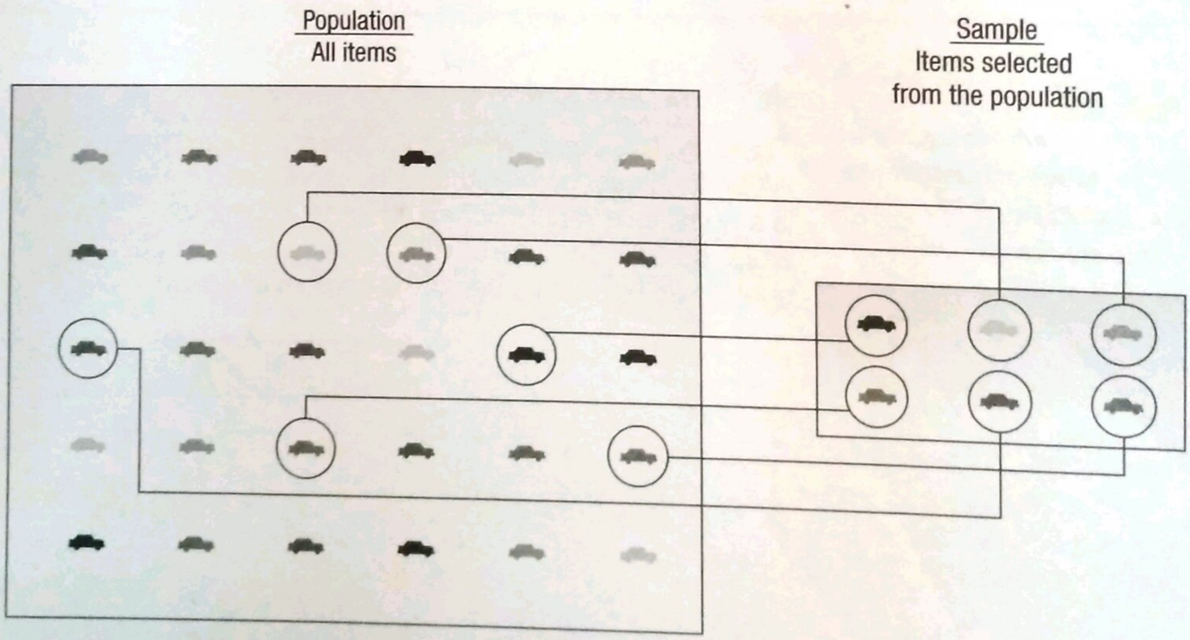
Sometimes we must make decisions based on a limited set of data. For example, we would like to know the operating characteristics, such as fuel efficiency measured by miles per gallon, of sport utility vehicles (SUVs) currently in use. If we spent a lot of time, money, and effort, all the owners of SUVs could be surveyed. In this case, our goal would be to survey the **population** of SUV owners.

> **POPULATION** The entire set of individuals or objects of interest or the measurements obtained from all individuals or objects of interest.

However, based on inferential statistics, we can survey a limited number of SUV owners and collect a **sample** from the population.

> **SAMPLE** A portion, or part, of the population of interest.

Samples are often used to obtain reliable estimates of population parameters. (Sampling is discussed in Chapter 8). In the process, we make trade-offs between the time, money, and effort to collect the data and the error of estimating a population parameter. The process of sampling SUVs is illustrated in the following graphic. In this example, we would like to know the mean or average SUV fuel efficiency. To estimate the mean of the population, six SUVs are sampled and the mean of their MPG is calculated.



So, the sample of six SUVs represents evidence from the population that we use to reach an inference or conclusion about the average MPG for all SUVs. The process of sampling from a population with the objective of estimating properties of a population is called **inferential statistics.**

> **INFERENTIAL STATISTICS** The methods used to estimate a property of a population on the basis of a sample.

Inferential statistics is widely applied to learn something about a population in business, agriculture, politics, and government, as shown in the following examples:

- Television networks constantly monitor the popularity of their programs by hiring Nielsen and other organizations to sample the preferences of TV viewers. For example, 10.5% of a sample of households with TVs watched *The Big Bang Theory* during the week of February 25, 2012 (www.nielsen.com). These program ratings are used to make decisions about advertising rates and whether to continue or cancel a program.
- In 2012, a sample of 40 U.S. Internal Revenue Service volunteer program sites was selected and the volunteer tax preparers were tested with three standard tax returns. The sample indicated that tax returns were completed with a 49% accuracy rate. In this example, the statistics are used to make decisions about how to improve the accuracy rate by correcting the most common errors and improving the training of volunteers. (www.treasury.gov/tigta/auditreports/2012reports/201240088fr.pdf)

A feature of our text is self-review problems. There are a number of them interspersed throughout each chapter. The first self-review follows. Each self-review tests your comprehension of preceding material. The answer and method of solution are given in Appendix E. You can find the answer to the following self-review in 1–1 in Appendix E. We recommend that you solve each one and then check your answer.

SELF-REVIEW 1–1

The answers are in Appendix E.

The Atlanta-based advertising firm Brandon and Associates asked a sample of 1,960 consumers to try a newly developed chicken dinner by Boston Market. Of the 1,960 sampled, 1,176 said they would purchase the dinner if it is marketed.

(a) What could Brandon and Associates report to Boston Market regarding acceptance of the chicken dinner in the population?
(b) Is this an example of descriptive statistics or inferential statistics? Explain.

LO1-4
Classify variables as qualitative or quantitative, and discrete or continuous.

TYPES OF VARIABLES

There are two basic types of variables: (1) qualitative and (2) quantitative (see Chart 1–2). When an object or individual is observed and recorded as a nonnumeric characteristic, it is a qualitative variable or an attribute. Examples of qualitative variables are gender, beverage preference, type of vehicle owned, state of birth, and eye color. When a variable is qualitative, we usually count the number of observations for each category and determine

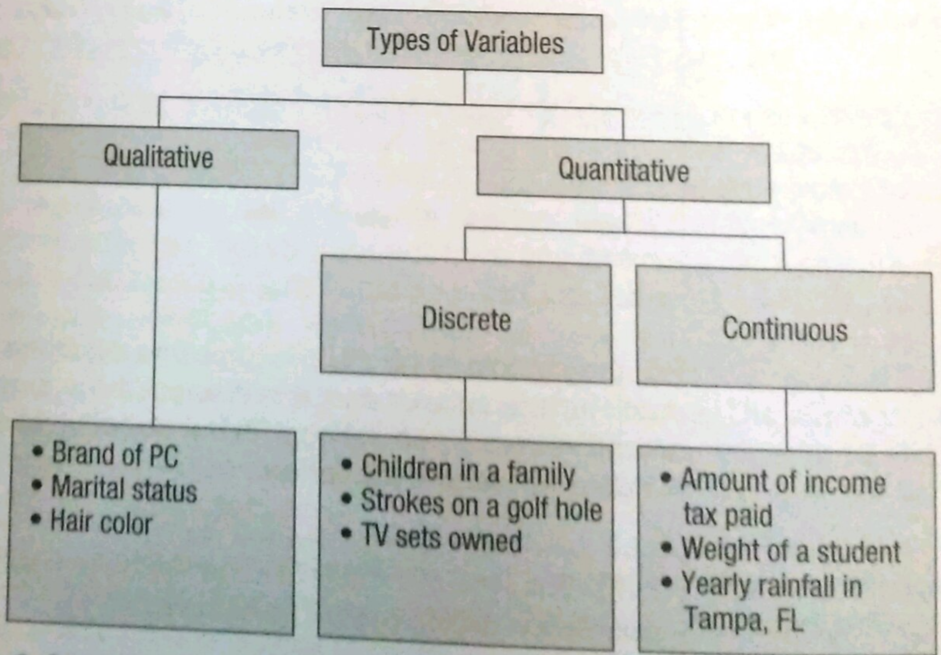


CHART 1–2 Summary of the Types of Variables

what percent are in each category. For example, if we observe the variable eye color, what percent of the population has blue eyes and what percent has brown eyes? If the variable is type of vehicle, what percent of the total number of cars sold last month were SUVs? Qualitative variables are often summarized in charts and bar graphs (Chapter 2).

When a variable can be reported numerically, the variable is called a quantitative variable. Examples of quantitative variables are the balance in your checking account, the ages of company presidents, the life of a car battery (such as 42 months), and the number of people employed by a company.

Quantitative variables are either discrete or continuous. Discrete variables can assume only certain values, and there are "gaps" between the values. Examples of discrete variables are the number of bedrooms in a house (1, 2, 3, 4, etc.), the number of cars arriving at Exit 25 on I-4 in Florida near Walt Disney World in an hour (326, 421, etc.), and the number of students in each section of a statistics course (25 in section A, 42 in section B, and 18 in section C). We count, for example, the number of cars arriving at Exit 25 on I-4, and we count the number of statistics students in each section. Notice that a home can have 3 or 4 bedrooms, but it cannot have 3.56 bedrooms. Thus, there is a "gap" between possible values. Typically, discrete variables are counted.

Observations of a continuous variable can assume any value within a specific range. Examples of continuous variables are the air pressure in a tire and the weight of a shipment of tomatoes. Other examples are the ounces of raisins in a box of raisin bran cereal and the duration of flights from Orlando to San Diego. Grade point average (GPA) is a continuous variable. We could report the GPA of a particular student as 3.2576952. The usual practice is to round to 3 places—3.258. Typically, continuous variables result from measuring.

LO1-5
Distinguish between nominal, ordinal, interval, and ratio levels of measurement.

LEVELS OF MEASUREMENT

Data can be classified according to levels of measurement. The level of measurement determines how data should be summarized and presented. It will also indicate the type of statistical analysis that can be performed. Here are two examples of the relationship between measurement and how we apply statistics. There are six colors of candies in a bag of M&Ms. Suppose we assign brown a value of 1, yellow 2, blue 3, orange 4, green 5, and red 6. What kind of variable is the color of an M&M? It is a qualitative variable. Suppose someone summarizes M&M color by adding the assigned color values, divides the sum by the number of M&Ms, and reports that the mean color is 3.56. How do we interpret this statistic? You are correct in concluding that it has no meaning as a measure of M&M color. As a qualitative variable, we can only report the count and percentage of each color in a bag of M&Ms. As a second example, in a high school track meet there are eight competitors in the 400-meter run. We report the order of finish and that the mean finish is 4.5. What does the mean finish tell us? Nothing! In both of these instances, we have not used the appropriate statistics for the level of measurement.

There are four levels of measurement: nominal, ordinal, interval, and ratio. The lowest, or the most primitive, measurement is the nominal level. The highest is the ratio level of measurement.

Nominal-Level Data

For the **nominal level of measurement,** observations of a qualitative variable are measured and recorded as labels or names. The labels or names can only be classified and counted. There is no particular order to the labels.

NOMINAL LEVEL OF MEASUREMENT Data recorded at the nominal level of measurement is represented as labels or names. They have no order. They can only be classified and counted.

The classification of the six colors of M&M milk chocolate candies is an example of the nominal level of measurement. We simply classify the candies by color. There is no natural order. That is, we could report the brown candies first, the orange first, or any of the other colors first. Recording the variable gender is another example of the nominal level of measurement. Suppose we count the number of students entering a football game with a student ID and report how many are men and how many are women. We could report either the men or the women first. For the data measured at the nominal level, we are limited to counting the number in each category of the variable. Often, we convert these counts to percentages. For example, a study of the color of M&M candies reports the following results (www.sensationalcolor.com/color-trends/most-popular-colors-177/mam-colors.html):

Color	Percent in a bag
Blue	24%
Green	20
Orange	16
Yellow	14
Red	13
Brown	13

STATISTICS IN ACTION

Where did statistics get its start? In 1662 John Graunt published an article called "Natural and Political Observations Made upon Bills of Mortality." The author's "observations" were the result of a study and analysis of a weekly church publication called "Bill of Mortality," which listed births, christenings, and deaths and their causes. Graunt realized that the Bills of Mortality represented only a fraction of all births and deaths in London. However, he used the data to reach broad conclusions about the impact of disease, such as the plague, on the general population. His logic is an example of statistical inference. His analysis and interpretation of the data are thought to mark the start of statistics.

To process the data for a variable measured at the nominal level, we often numerically code the labels or names. For example, if we are interested in measuring the home state for students at East Carolina University, we would assign a student's home state of Alabama a code of 1, Alaska a code of 2, Arizona a 3, and so on. Using this procedure with an alphabetical listing of states, Wisconsin is coded 49 and Wyoming 50. Realize that the number assigned to each state is still a label or name. The reason we assign numerical codes is to facilitate counting the number of students from each state with statistical software. Note that assigning numbers to the states does not give us license to manipulate the codes as numerical information. Specifically, in this example, $1 + 2 = 3$ corresponds to Alabama + Alaska = Arizona. Clearly, the nominal level of measurement does not permit any mathematical operation that has any valid interpretation.

Ordinal-Level Data

The next higher level of measurement is the **ordinal level.** For this level of measurement a qualitative variable or attribute is either ranked or rated on a relative scale.

> **ORDINAL LEVEL OF MEASUREMENT** Data recorded at the ordinal level of measurement is based on a relative ranking or rating of items based on a defined attribute or qualitative variable. Variables based on this level of measurement are only ranked or counted.

Best Business Climate
1. Texas
2. Utah
3. Virginia
4. Florida
5. Louisiana
6. Indiana
7. South Carolina
8. Tennessee
9. Georgia
10. Nebraska

For example, many businesses make decisions about where to locate their facilities; in other words, where is the best place for their business? Business Facilities (www.businessfacilities.com) publishes a list of the top 10 states for the "best business climate." The 2012 rankings are shown to the left. They are based on the evaluation of 20 different factors, including the cost of labor, business tax climate, quality of life, transportation infrastructure, educated workforce, and economic growth potential to rank states based on the attribute "best business climate."

This is an example of an ordinal scale because the states are ranked in order of best to worst business climate. That is, we know the relative order of the states based on the attribute. For example, in 2012 Texas had the best business climate. Louisiana was fifth, and that was better than South Carolina but not as good as Virginia. Notice that we cannot say that Texas's business climate is five times better than Louisiana's business climate because the magnitude of the differences between the states is not known.

Another example of the ordinal level measure is based on a scale that measures an attribute. This type of scale is used when students rate instructors on a variety of attributes. One attribute may be: "Overall, how do you rate the quality of instruction in this class?" A student's response is recorded on a relative scale of inferior, poor, good, excellent, and superior. An important characteristic of using a relative measurement scale is that we cannot distinguish the magnitude of the differences between groups. We do not know if the difference between "Superior" and "Good" is the same as the difference between "Poor" and "Inferior."

Table 1–1 lists the frequencies of student ratings of instructional quality for Professor James Brunner in an Introduction to Finance course. The data are summarized based on the order of the scale used to rate the instructor. That is, they are summarized by the number of students who indicated a rating of superior (6), good (28), and so on. We can also convert the frequencies to percentages. About 37.8% of the students rated the instructor as good.

TABLE 1–1 Rating of a Finance Professor

Rating	Frequency
Superior	6
Good	28
Average	25
Poor	12
Inferior	3

Interval-Level Data

The **interval level of measurement** is the next highest level. It includes all the characteristics of the ordinal level, but, in addition, the difference or interval between values is meaningful.

> **INTERVAL LEVEL OF MEASUREMENT** For data recorded at the interval level of measurement, the interval or the distance between values is meaningful. The interval level of measurement is based on a scale with a known unit of measurement.

The Fahrenheit temperature scale is an example of the interval level of measurement. Suppose the high temperatures on three consecutive winter days in Boston are 28, 31, and 20 degrees Fahrenheit. These temperatures can be easily ranked, but we can also determine the interval or distance between temperatures. This is possible because 1 degree Fahrenheit represents a constant unit of measurement. That is, the distance between 10 and 15 degrees Fahrenheit is 5 degrees, and is the same as the 5-degree distance between 50 and 55 degrees Fahrenheit. It is also important to note that 0 is just a point on the scale. It does not represent the absence of the condition. The measurement of zero degrees Fahrenheit does not represent the absence of heat or cold. But by our own measurement scale, it is cold! A major limitation of a variable measured at the interval level is that we cannot make statements similar to 20 degrees Fahrenheit is twice as warm as 10 degrees Fahrenheit.

Another example of the interval scale of measurement is women's dress sizes. Listed below is information on several dimensions of a standard U.S. woman's dress.

Size	Bust (in)	Waist (in)	Hips (in)
8	32	24	35
10	34	26	37
12	36	28	39
14	38	30	41
16	40	32	43
18	42	34	45
20	44	36	47
22	46	38	49
24	48	40	51
26	50	42	53
28	52	44	55

Why is the "size" scale an interval measurement? Observe that as the size changes by two units (say from size 10 to size 12 or from size 24 to size 26), each of the measurements increases by 2 inches. To put it another way, the intervals are the same.

There is no natural zero point for dress size. A "size 0" dress does not have "zero" material. Instead, it would have a 24-inch bust, 16-inch waist, and 27-inch hips. Moreover, the ratios are not reasonable. If you divide a size 28 by a size 14, you do not get the same answer as dividing a size 20 by a size 10. Neither ratio is equal to two, as the "size" number would suggest. In short, if the distances between the numbers make sense, but the ratios do not, then you have an interval scale of measurement.

Ratio-Level Data

Almost all quantitative variables are recorded on the ratio level of measurement. The ratio level is the "highest" level of measurement. It has all the characteristics of the interval level, but, in addition, the 0 point and the ratio between two numbers are both meaningful.

RATIO LEVEL OF MEASUREMENT Data recorded at the ratio level of measurement are based on a scale with a known unit of measurement and a meaningful interpretation of zero on the scale.

Examples of the ratio scale of measurement include wages, units of production, weight, changes in stock prices, distance between branch offices, and height. Money is also a good illustration. If you have zero dollars, then you have no money, and a wage of $50 per hour is two times the wage of $25 per hour. Weight is also measured at the ratio level of measurement. If a scale is correctly calibrated, then it will read 0 when nothing is on the scale. Further, something that weighs 1 pound is half as heavy as something that weighs 2 pounds.

Table 1–2 illustrates the ratio scale of measurement for the variable, annual income for four father-and-son combinations. Observe that the senior Lahey earns twice as much as his son. In the Rho family, the son makes twice as much as the father.

TABLE 1–2 Father–Son Income Combinations

Name	Father	Son
Lahey	$80,000	$ 40,000
Nale	90,000	30,000
Rho	60,000	120,000
Steele	75,000	130,000