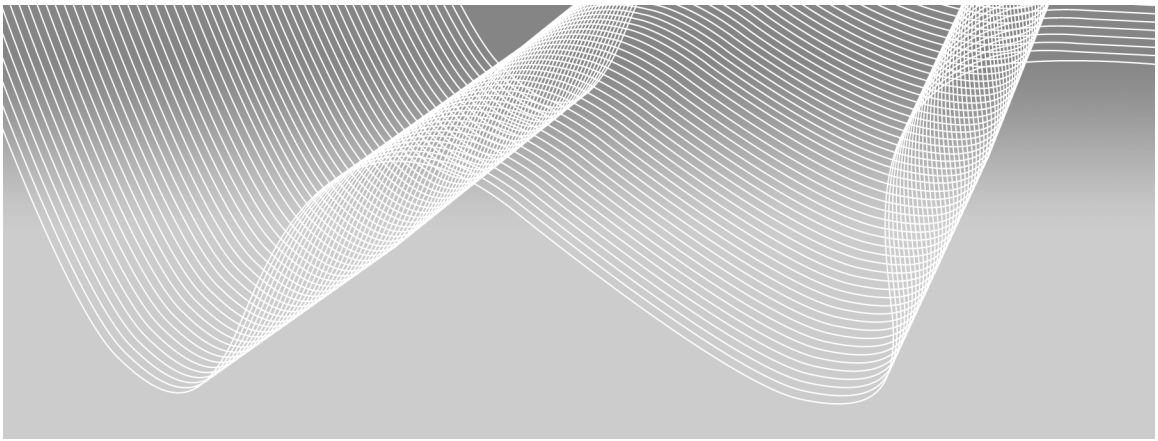


Chapter 1

About Data

1.1 Introduction

Data are an essential ingredient in learning about the business world and solving business problems. The list that follows gives you a flavor of the numerous business situations requiring data:

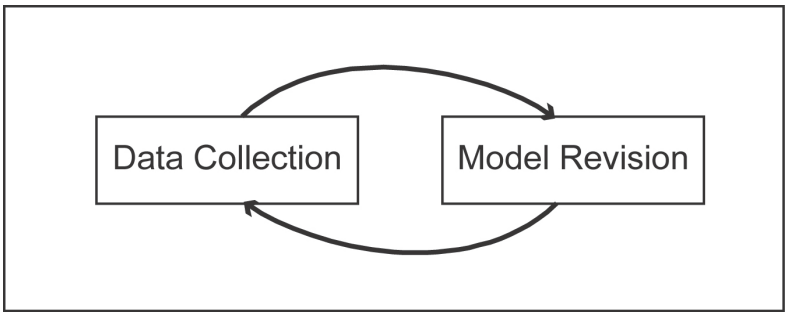
- In monitoring the delinquency rate of credit card accounts, a bank needs data on delinquencies and the demographic factors that might have contributed to them.
- In evaluating the performance of bank branches, a bank needs data on the branches and the business environment under which each operates.
- In making investment choices, a fund manager needs data to assess the risks associated with investment alternatives.
- In developing a new test for blood sugar levels of diabetics, a company needs to estimate the size of the market from various data sources.
- In making an advertising claim that a dishwasher is superior to functionally equivalent models of several competitors, a company needs to collect data that substantiate that claim.
- In comparing the performance of several suppliers of a plastic material, a car manufacturer needs data on the strength of each supplier's material.
- In new product development, a company requires data on customers' needs and desires.
- In developing an ink for printing on plastic bags, a company needs experimental data that evaluate the adhesiveness of various ink mixtures.
- In pricing a multi-year warranty program for personal computers, a computer manufacturer needs data on the failure rate of components and the cost of replacing them.
- In getting a new cholesterol drug ready for approval, a company needs data that prove that the new drug works and has acceptable risks associated with it.
- In designing a new wastewater treatment facility, a company needs data from a simulation to evaluate the performance of various plant configurations.

Data are measurements of facts that are closely linked to statistical models. Models provide explanations in view of the data. Statistical models are often mathematical expressions. Data are used to estimate model coefficients or parameters of the mathematical expressions. However, the process of statistical model building goes further. Data and statistics help decide which variables to include in the model, as well as how to express variables in combination with other variables.

For example, in building a model for the valuation of medium-size businesses, you might use data to decide whether to include total sales or total assets or both as model variables. Data also help determine whether the value of a business depends linearly on sales and assets or whether a change of scale yields a better model. Last, additional data often suggest modifications to existing models, such as adding new variables, eliminating existing ones, or re-estimating the model coefficients.

Data suggest modifications to existing empirical models. The cycle of data collection and model building reflects the fact that businesses evolve dynamically. They require continuous updating of facts and assumptions as well as continuous adaptation of the model to new realities as suggested by the data. This empirical learning cycle is shown in Figure 1.1.

Figure 1.1 Empirical Learning Cycle



This chapter examines some common sources of data: existing data sources, survey data, data from designed experiments, observational data, and data from simulation experiments. You should recognize from the outset that data reflect reality but are not reality itself. Even though data often seem to give a true picture, there are occasions when data present an incomplete and distorted picture of reality. Whenever data are used, examine the quality and relevance of the data for the purpose to which they are put.

Data are generated or obtained from sources either internal or external to the business. Data might already exist or have to be collected. Sometimes a mixture of internal and external data needs to be used. For example, in benchmarking order processing, a business might compare its own order process to a similar process within the same company, to those of (external) competitors, or even to those of noncompetitors.

1.2 Why Data Are Needed

Management requires data because they encounter variation and uncertainty in decision making. Especially in larger businesses, management has an undiminished need to understand the business and its environment. Global markets have added further complexity and the need to adjust decisions quickly to changes in the business environment. Data provide understanding of the internal business operations as well as of the external environment.

The saying "You cannot manage what you cannot measure!" expresses how essential data are to modern management practice. Data guide strategic, tactical, and operational decision making. Management needs data to determine the future direction of a business, to allocate resources, and to run existing operations smoothly and efficiently. A diversified company needs data to determine which businesses to foster and grow, which to acquire, and which to leave behind. The data for such strategic decisions are not easy to come by. They require the proper context to be useful and informative. A retail chain needs data to determine sites for new retail outlets. Retailers maintain and acquire large databases for such tactical decisions. These data are fed into specialized software to evaluate locations for their suitability. Such software determines the data needs.

Management needs data because business operations, markets, and finance, among others, cannot be predicted with certainty. The outcomes of virtually all business processes are subject to some variation. Markets tend to shift randomly over the short term, often hiding underlying long-term patterns. Therefore, a major reason behind the need for data is to understand variation.

Data play an essential role in managing a business. Data are used in product development, manufacturing, and marketing as well as in support functions such as finance and human resources. Data are needed to

- **demonstrate compliance** with the regulatory requirements of such agencies as the Environmental Protection Agency
- **explain** relationships between variables
- **estimate** a curvilinear relationship that pinpoints a staffing level above which processing times do not increase
- **develop** new products and services by providing quantitative and qualitative information about customer preferences
- **test** the product of one company against similar ones of competitors
- **monitor** the quality of products and customer services
- **predict** the market for various products or services

Data alone would be insufficient for successful management. Data are an aid in decision making. They do not replace business knowledge but increase it. Subject matter knowledge gives management direction in several ways. It guides data collection and is an essential prerequisite for proper data analysis and model building.

Data are needed because of variation. Without variation in customer opinion, one customer would be sufficient to find out what all other customers like and dislike about a product or service. However, this case of customer unanimity is hard to find. Variation occurs when a process yields a finite or infinite variety of uncertain outcomes such as the daily changes in the stock market, varying times to complete a task, or varying dimensions of a part used in the manufacture of a product. Repeated observations of outcomes of the same process under nearly identical conditions yield different values. Process variation is a major cause of concern in improving the quality of products and services. Eliminating variation, or at least taming it by explaining, mitigating, and eliminating some of its causes, is a major effort in continuous improvement efforts. Variation, especially that arising from the uncertainty of the future, can never be eliminated.

1.3 Sources of Data

Data need to be assembled, acquired, or collected. Because data represent factual information, data quality is an important issue. Unfortunately, no matter how they were collected, data often contain errors. In this section, various sources of data and their most common advantages and disadvantages are discussed.

1.3.1 Existing Data versus New Data

The data required to proceed to solve a specific problem might already exist within a business or they might be obtained from commercial and noncommercial sources outside the business. The main task with such existing data is to locate them, verify their quality, and use them judiciously.

Sometimes, existing data are used instead of collecting new data. For example, in estimating the beta risk of a company, you need to rely on recent market prices of a company's stock. Other times, newly collected data are more appropriate. For example, an Internet service provider might contemplate a variety of new services to a market segment. In order to find out which services customers are most likely to accept, new data need to be collected because none exist. New data can be tailored to solve a particular problem, whereas existing data are limited by the available content. We now briefly discuss some data sources that play an important role in business applications.

1.3.2 Existing Data

The first task with existing data is to find the appropriate ones. Sources are varied and often unreliable. When they are the only data available, they might provide better value than trying to obtain new data or not using data at all.

Sources

Sources for existing data are plentiful. Within a company, for example, a business unit providing computer services for databases and Web applications logs data continuously about the usage and update frequency of its servers. Such information is helpful in improving the operation of the servers.

Data might exist but in an unusable format requiring work to make them useful. Good data require some effort to assemble and verify. They usually require considerable resources but often prove themselves as good investments. Examples of existing business internal data are

- delivery times of packages of a package delivery business
- monthly sales by department for the past 10 years of a chain store business
- cost of transplant surgery in a hospital for the most recent 3 years
- results of strength tests of a material for the last year of a material testing lab
- monthly deposits at a bank branch of a bank holding company

Many organizations provide financial or marketing data by subscription or on an individual basis. Benchmarking organizations offer inter- and intra-industry data on a variety of subjects. Financial data on companies are available from organizations such as COMPUSTAT, which is a subsidiary of Standard & Poor's. Another popular financial database is from the Center for Research in Security Prices (CRSP). CRSP is an organization with strong ties to academic researchers. For financial market data, finance.yahoo.com is used in this book. Marketing data are available from many organizations; among the best known is ACNielsen. Specialist firms such as Claritas.com provide data and perform studies on consumer spending, product usage, lifestyle segmentation, and many other applications.

Uses of Existing Data

Existing data can be summarized and analyzed for empirical patterns or relationships that hold true for the past. Existing data are most useful in

- setting standards or targets for formulating future hypotheses
- comparing present or new practices with past practices

Existing data are sometimes used to construct models for predicting the future. In this case, you should use caution because the historical pattern might not predict the future. In using existing data, consider the following points:

- Verify the integrity of the data source. Many data sources, including those that are commercially sold, contain errors. In order to avoid erroneous analyses, always include sanity checks in your preliminary analysis.
- Newly accruing data might change a distribution's shape and the appearance of graphs and numerical summaries.
- An important question in business applications is how long a data history should be considered. Going too far back into the historical record could present information that has little relevance to the current problem. Using a very short history might not reveal important patterns.
- Historical data are often not random and, thus, limit the objectivity of the conclusions you can draw from them.
- Different methods of presentation (for example, histograms) might give different impressions of the data. Use JMP to try many methods and see which one is most useful (see Chapter 3).

1.3.3 New Data

The advantage of new data is that they can be collected with a focus on obtaining problem-specific answers. By using problem-specific variables and optimum data collection methods, you can get results that provide answers that are more specific. However, new data often require considerable expenditures of time and resources and, thus, present the need to trade off the benefits of special-purpose data against the cost and time of acquiring them. Different methods of data collection might be appropriate. Focus groups provide general and highly qualitative insights. Surveys, designed experiments, and simulation often provide very specific quantitative and qualitative data to answer specific questions.

Focus Groups

Focus groups are structured discussion groups of individuals with the aim of obtaining several perspectives on a single problem. Focus groups are useful for exploring personal attitudes and beliefs as well as experiences and reactions from individuals in a group setting. Businesses and politicians use them at an early stage of the problem-solving process. Focus group interviews provide qualitative information. Findings are open-ended and not constrained by finite choices. As a result, they can help generate ideas and develop questions for a follow-up questionnaire. A topic for a focus group might be to investigate the importance of certain features on washing machines. The moderator might give a broad outline of washing machine types (top-loading versus front-loading) and then try to elicit those features that are important to the group members. The data

obtained from focus groups often lend themselves to qualitative, rather than quantitative, analysis.

Focus group sessions are administered by a skilled focus group moderator who is neutral, skilled in leading a group, and has good interpersonal skills. Focus group interviews require more planning than other types of interviews. Participants need to meet at specific times and places that are especially equipped for recording the findings. The recommended size for focus groups is 6 to 10 participants per group, although there have been groups as large as 20. Focus group sessions usually last 1 or 2 hours.

Here are some important things to remember:

- Focus groups often help us understand why people attach importance to an issue or a product feature.
- Focus group findings should not be generalized to the entire population without further verification.
- Focus group moderators must not influence the group.
- Focus groups yield results that are open-ended in nature.

Surveys

Surveys consist of a planned and designed collection, analysis, and interpretation of data regarding some aspect of a well-defined population. Populations could be people living in a particular region or having specific characteristics or some other identification of interest to the organization conducting the survey. In business, it is common to conduct satisfaction surveys on past and existing customers. For example, a bank might conduct a quarterly survey of new customers with the aim of monitoring their satisfaction with bank services. Surveys are regularly conducted in certain areas of economic activity, such as housing construction, manufacturing, or household expenditures.

Census

Surveys can collect data from all members of the designated population. In that case, they are called a *census*. The best-known census, prescribed in the U.S. Constitution, is the one conducted every 10 years by the U.S. Census Bureau that covers all people living in the United States. Censuses are also conducted on many smaller subpopulations, such as all the employees of a plant site who are asked about their health care preferences. Censuses of large populations are problematic because it is not always possible to reach all parts of the populations. The missing responses might lead to biased conclusions.

A 100% census can be considered when the cost of sampling is negligible or the population is relatively small and the precision with which results have to be ascertained is fairly high. In stratified audit samples of sales taxes paid, the population is separated into non-overlapping strata by merchandise amount. The stratum containing the high value items often is subject to a census, whereas lower valued strata are sampled.

Sample Surveys

While a census attempts to survey 100% of the members of the designated population, a sample survey aims at surveying only a subpopulation. Depending on how the sample subpopulation is selected, it can be called a *representative sample*, *random sample*, or *judgment sample*, for example. These terms are explained in the next chapter.

In comparing the merits of a sample survey versus a census, several features stand out. Table 1.1 gives a brief comparison of sample survey and census.

Table 1.1 Comparison of Sample and 100% Census

Sample	100% Census
Inexpensive	Expensive
Possible	Often impossible
Allows a more intensive look	Allows a superficial look
Unbiased due to random selection	Biased due to often small number of non-responses

Sample surveys are often cheaper to conduct because fewer population members have to be surveyed. When a large or geographically dispersed population needs to be surveyed, sample surveys can produce results in a more timely fashion than censuses. In a sample survey, you can trade some of the savings due to having to examine fewer population members with a more in-depth look at each member, which also might lead to higher quality data. Last, looking at 100% of a population does not guarantee an unbiased picture, especially when reality shows that some subpopulations might be nearly impossible to examine. Sample surveys carry with them potential disadvantages. Even random selection does not guarantee representativeness. The number of population members surveyed in the sample, the sample size, might be inadequate to draw the desired conclusions.

Designed Experiments

Design of experiments (DOE) is a methodology of systematically varying the inputs, or X-factors, to measure their effect on the output, the Y-response variable, under well-defined experimental conditions. Experiments involve active manipulation of X-factors to study their effects on Y. DOE methodology tells us which factor level combinations to include in the experiment. Consequently, factor effects are measured precisely, accurately, and efficiently. In DOE, you can conduct smaller experiments and augment them by additional experiments. This helps to improve the results, reconcile ambiguities, and bring the experimenter nearer to a cumulative understanding of the problem.

Several points are important with experimental data. Experiments are often performed in laboratories and on a small scale. One needs to be careful when extrapolating results to a larger scale. Experiments are often performed under well-controlled conditions. Extrapolation of controlled conditions to field conditions is often dangerous, because of the additional sources of variation present in the field. Experiments have the ability to manipulate factors (causes) systematically and yield insights unobtainable from historical or field data.

There are many areas where experimentation is impossible or severely restricted (for example, because of ethical concerns).

Observational Studies

Observational studies are occasionally referred to as *quasi-experimental studies* because data are collected using factor combinations similar to those in DOE, but without the benefit of random allocations. Factor level allocations are non-random, because the experimenter cannot manipulate factor level allocations at will. The experimenter may be able to match subject characteristics against desired factor allocations. For example, in studying consumer car buying behavior, personal income is an important X-factor. A subject is selected on the basis of income level. However, the experimenter cannot ask a subject to make more or less money, but can only match subjects to factor levels of their existing income. Because of the difficulty in matching complex factor-level patterns, the number of factors in an observational study is limited.

Computer Simulation

Simulation is a useful tool to understand the behavior of complex systems. With *simulation*, systems can be observed even outside the safe range of operations. Design changes can be evaluated even before the changes are actually constructed. Computer simulation allows the creation of very realistic (although still artificial) models that often match historical data. Computer simulation models also allow for a manipulation of factors in the model so that they can be evaluated regarding their importance on the output variable.

A simulation model is only as good as the assumptions on which it is based. Different people often disagree over those assumptions and the results. The controversy regarding predictions from simulation models concerning climate change is an example. Different experimenters make different assumptions and so predict a different future.

1.4 Data Scales

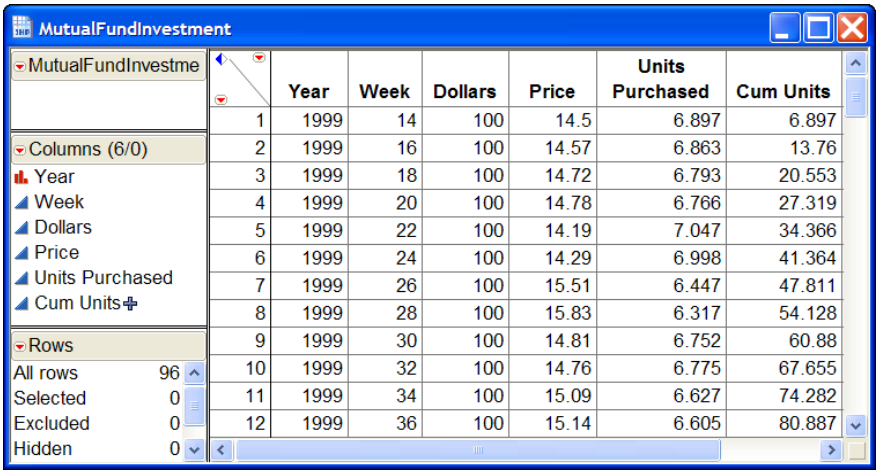
Data are recorded observations about the physical or perceptual world. In the physical sciences, these data often represent measurements on physical objects. These might be measurements of naturally occurring phenomena or the result of carefully designed experiments. In business, data often occur quite regularly. For example, they occur as regular records of daily sales volume or as information specifically gathered through surveys and experiments.

Data are characteristics measured on elements. *Elements* can be physical objects like customers or parts of a product. They can also be organizationally complex entities such as bank branches, business divisions, or entire businesses. A characteristic that can take on more than one value is called a *variable*. Otherwise, it is called a *constant*. In passenger cars, for example, the number of cylinders in the engine is a variable because it can take on values like 2, 4, 5, 6, 8, 12, and even 16. The number of axles is constant at 2, excepting some special designs that might have more than 2.

In JMP as in other statistical software, data are represented in tables. Columns represent variables or characteristics. Rows represent the elements on which these characteristics have been measured. Figure 1.2 shows an excerpt of the JMP data file *MutualFundInvestment.jmp* containing a record of mutual fund payroll deductions. The following variables are in six columns:

- Column 1: Year (only 1999 is shown)
- Column 2: Week (weeks 14 to 36 are shown)
- Column 3: Dollars (amount invested per week)
- Column 4: Price (per unit of mutual fund)
- Column 5: Units Purchased (calculated from Columns 3 and 4)
- Column 6: Cum Units (cumulative units calculated from Column 5)

Figure 1.2 Data File Excerpt for Mutual Fund Payroll Deduction



MutualFundInvestment

MutualFundInvestme

Columns (6/0)
- Year
- Week
- Dollars
- Price
- Units Purchased
- Cum Units

Rows
- All rows 96
- Selected 0
- Excluded 0
- Hidden 0

	Year	Week	Dollars	Price	Units Purchased	Cum Units
1	1999	14	100	14.5	6.897	6.897
2	1999	16	100	14.57	6.863	13.76
3	1999	18	100	14.72	6.793	20.553
4	1999	20	100	14.78	6.766	27.319
5	1999	22	100	14.19	7.047	34.366
6	1999	24	100	14.29	6.998	41.364
7	1999	26	100	15.51	6.447	47.811
8	1999	28	100	15.83	6.317	54.128
9	1999	30	100	14.81	6.752	60.88
10	1999	32	100	14.76	6.775	67.655
11	1999	34	100	15.09	6.627	74.282
12	1999	36	100	15.14	6.605	80.887

Each row of the data table represents a purchase action of mutual fund units. The weeks are ordered from earliest (Year 1999, Week 14) to latest (Year 2002, Week 46). The payroll deduction amount remained constant at $100. However, there are other amounts at year-end representing dividend disbursements.

The variable Price is measured on a continuous scale because any value greater than zero makes sense. Is $14.5 a high or a low price? That depends entirely on the comparison with other prices within the time frame of the mutual fund. Since the data file is from Week 14 in 1999 to Week 46 in 2002, use this period to judge a high or a low unit price. Suppose that a unit price of $10 or more is considered high and less than $10 is low. The variable Price is now simplified into two categories, high and low. Such a simplification of continuous variables often communicates concepts more effectively; at other times, they are the only form in which data are available. A classification of the unit price of a mutual fund into high and low is a different type of scale.

Numeric variables are often, but not always, measurements on a continuous scale. Such data are usually on the ratio scale or the interval scale. Variables might also record rankings of performance. These data are said to be on an ordinal scale. Last, variables might represent simple categories, such as left and right. These are on the nominal scale.

1.4.1 Ratio Scale

Ratio scales are the highest level of scales. They are called *ratio scales* because when two measurements differ by a multiplicative factor *r*, then the larger value is said to be *r* times larger than the smaller. For example, if the price of a barrel of oil is $100 and subsequently increases to $110, you can say that the price increased by 10%. The ratio factor is r=1.1. Likewise, a rise from $100 to $200 would represent a doubling in price, because twice the monetary exchange units are required to buy a barrel of oil. In this case, the ratio factor is r=2. Ratio scales have a natural 0. If crude oil costs $0, then no money is needed to buy a barrel of it.

1.4.2 Interval Scale

Interval scales are so called because when two measurements differ by an additive factor d, then the larger is said to be d units higher than the lower. For example, if the temperature in London is 50 degrees and the temperature in New York is 40 degrees, then London is 10 degrees warmer, but you cannot and should not say that London is 25% warmer (as for data on a ratio scale). The difference between the price of oil and the temperature in degrees Fahrenheit is that the Fahrenheit scale has an arbitrary 0 (at approximately –18 degrees Celsius), just as the Celsius scale has an arbitrary 0 at 32 degrees F.

1.4.3 Ordinal Scale

Data on an ordinal scale rank performance measurements. For example, an issue of *Consumer Reports* ranked DVD players from number 1 to number 13, suggesting that number 1 was the best performer and number 13 was the poorest performer of the 13 investigated. This scale did not suggest, however, that the number 1 DVD player was twice as good as the number 2 player nor that the difference between the player ranked 1 and the player ranked 2 was the same as the difference between players ranked 2 and 3. Other examples of variables with ordinal scales involve Likert scale responses discussed in Section 1.4.5.

1.4.4 Nominal Scale

Data on a nominal scale represent a set of categories that differ in some characteristic that does not necessarily have a measurable magnitude. Observations on a nominal scale are classified in one of several non-overlapping categories, also called the levels of the variables. Gender, Type of Bank, and Country of Origin are on a nominal scale.

Table 1.2 compares these four scales, relates them to JMP terminology, and identifies whether a mean or a standard deviation make sense. In JMP, the modeling type of each variable has a specific marker next to its name in the data window.

You can always collapse a higher order scale into a lower scale. For example, categorize days above 80 degrees Fahrenheit as hot, days between 40 and 80 degrees as temperate, and days below 40 degrees Fahrenheit as cold. The collapse is arbitrary because other temperature cutoff points could have been used.

Table 1.2 Measurement Scales

Scale	Modeling Type in JMP	Mean, Std. Dev.	Other Names
Ratio	Continuous	Yes	Quantitative (may be both continuous and discrete=integer)
Interval	Continuous	Yes	Quantitative (may be both continuous and discrete=integer)
Ordinal	Ordinal	No	Qualitative (cannot be continuous)
Nominal	Nominal	No	Qualitative (cannot be continuous)

1.4.5 Likert Scale

In survey questionnaires, the Likert scale is used for measuring attitudes in marketing research. Likert scales measure the strength of a respondent's perceived agreement or disagreement to statements such as "ABC bank employees are always friendly when doing business with their customers." A typical form of a response on a five-point Likert scale is

Strongly Disagree	Disagree	Neutral Opinion	Agree	Strongly Agree
1	2	3	4	5

In a five-point Likert scale, the responses are scored 1, 2, 3, 4, and 5. The value 3 is the neutral position, while 1 and 5 are the two extremes. The very common seven-point Likert scale augments the responses to Strongly Disagree (1), Disagree (2), Slightly Disagree (3), Neither Agree nor Disagree (4), Slightly Agree (5), Agree (6), and Strongly Agree (7). Here the neutral position is scored with 4. Likert scales are often treated as interval scales because you calculate average responses to statements. Other times they are treated as ordinal, such as when the ordering of the responses needs to be stressed.

1.5 Summary

- Data are needed because variation is unavoidable.
- Data collection and model revision are iterative approaches that are helpful in learning about and explaining business situations.
- In business, data come from many sources. Existing data are plentiful, but they may not always suit the purpose at hand. New data serve their purpose, but they require time and resources to collect and organize.
- Existing data need to be validated and checked for errors. Verify the assumptions under which they were collected. Make sure they describe situations that will continue into the future.
- The Internet is a rich source of data. Internet data need to be checked for quality.
- New data come from focus groups, surveys, designed experiments, observational studies, and computer simulation.
- Focus groups yield qualitative data that is difficult to generalize.
- Surveys are useful in collecting data about opinions and attitudes of large and widely dispersed populations.
- A census, although recording data on 100% of the population, does not necessarily provide better information than a well-selected sample. Censuses are useful for examining small populations or small subpopulations of larger populations.
- Samples, when collected following specific rules, yield unbiased information efficiently and with known precision.
- Designed experiments are useful when the X-factor levels can be easily manipulated. Designed experiments often yield results that are easy to interpret.
- Observational studies collect data in patterns similar to designed experiments. Subjects are chosen because their characteristics match desired X-factor level combinations.
- Data from computer simulations are used when it is impossible or impractical to identify realistic observations or to set-up realistic experiments.
- Data are usually organized by columns and rows. Columns represent variables or factors. Rows represent observations.
- In JMP, variables are continuous, ordinal, or nominal. Continuous variables allow calculation of means and standard deviations. Ordinal and nominal variables allow only counts of occurrences by nominal category.
- Surveys of attitudes often use Likert and similar scales. These scales, although ordinal or nominal in appearance, are treated as continuous in many applications.

1.6 Problems

1. Consider this question: "What is a good apple?" Identify variables that would address this question from several different perspectives:

   a. the apple grower's perspective

   b. the retailer's perspective

   c. the consumer's perspective. Discuss how these different perspectives influence the determination of the quality-defining characteristics.

2. Consider your daily trip to work or school:

   a. Describe the variability in your arrival time.

   b. Identify those factors that are in your control and those that are not within your control.

   c. Describe ways in which your travel process could be changed to reduce the variability.

3. A distribution center for a regional grocery chain is concerned about the losses they are incurring from spoilage among their produce items. This requires labor to inspect and remove spoiled items prior to shipment to a grocery store. In addition, the storage bins must be cleaned more frequently.

   a. Choose an appropriate performance variable and measurement scale for that variable.

   b. Identify important X-factors that characterize produce inventory. For each factor, identify the measurement scale selected (ratio, interval, ordinal, or nominal).

4. A human resources (HR) department is charged with evaluating worker satisfaction with the employee assistance program. The employee assistance program provides resources to help workers in the following areas: childcare, eldercare, care for disabled or chronically ill family members, financial planning, and funding for college.

   The HR department would like to collect data from their employees. Options for collecting data include a focus group, census, or sample survey. Prepare a one-page summary that discusses the advantages and disadvantages of each data collection method for this situation. Give your recommendation for the method that should be used. Provide reasons for your recommendation.

1.7 Case Study: Green's Gym—Part 1

Green's Gym is an independent, locally owned fitness center that serves a suburban area of approximately 65,000. A national fitness center chain is expanding into the region, threatening to reduce Green's market share. Green's offers individual, family, seasonal, and senior citizen memberships. Senior citizens receive a 15% discount. Seasonal memberships are either summer (May to August, targeting college students who have returned home for the summer) or winter (November to April, targeting those who prefer outdoor exercise during the warm months).

The facility includes a gymnasium, weight room, dance studio, fitness room (containing exercise equipment such as stationary bicycles and rowing machines), and a childcare facility. The following services are provided:

- Childcare for up to 25 children under the age of six while parents are using the facility. The childcare is not intended to provide daycare for working parents. Childcare hours are from 8 a.m. to 5 p.m. Monday–Friday.
- Adult evening volleyball and basketball leagues that meet once a week.
- The fitness center and weight room are open from 7 a.m. to 9 p.m., seven days a week.
- Saturday morning children's programs utilize the gym from 9 a.m. to 11 a.m. The balance of the weekend time is available for patron use.
- Aerobics classes are held in the dance studio during the week at 6:30 a.m., 10 a.m., and 7 p.m., and on Saturdays at 9 a.m. and 11 a.m. Senior citizen exercise classes are held on weekdays at 9 a.m. and 2 p.m.

Business Goals

Green's is interested in maintaining customer loyalty in the face of new competition.

Strategy

Green's would like to identify those factors that are important in retaining their members in order to prepare a customer satisfaction survey. The information collected in the survey will enable Green's to adjust their services to meet their customers' needs better, thereby retaining members.

Tasks

Perform the following tasks:

1. Identify three important factors that contribute to satisfied members at Green's Gym. Rank those factors in order of importance.

2. Discuss how each of these factors can be effectively measured. Recommend an appropriate measurement scale for each.
3. Are there any additional data that should be collected to help Green's assess their customer satisfaction?