CHAPTER 2

Sampling and Measurement

To analyze social phenomena with a statistical analysis, *descriptive* methods summarize the data and *inferential* methods use sample data to make predictions about populations. In gathering data, we must decide which subjects to sample. Selecting a sample that is representative of the population is a primary topic of this chapter.

Given a sample, we must convert our ideas about social phenomena into data through deciding what to measure and how to measure it. Developing ways to measure abstract concepts such as achievement, intelligence, and prejudice is one of the most challenging aspects of social research. A measure should have *validity*, describing what it is intended to measure and accurately reflecting the concept. It should also have *reliability*, being consistent in the sense that a subject will give the same response when asked again. Invalid or unreliable data-gathering instruments render statistical manipulations of the data meaningless.

The first section of this chapter introduces definitions pertaining to measurement, such as types of data. The other sections discuss ways, good and bad, of selecting the sample.

2.1 VARIABLES AND THEIR MEASUREMENT

Statistical methods help us determine the factors that explain *variability* among subjects. For instance, variation occurs from student to student in their college grade point average (GPA). What is responsible for that variability? The way those students vary in how much they study per week? in how much they watch TV per day? in their IQ? in their college board score? in their high school GPA?

Variables

Any characteristic we can measure for each subject is called a ***variable***. The name reflects that values of the characteristic *vary* among subjects.

Variable

A ***variable*** is a characteristic that can vary in value among subjects in a sample or population.

Different subjects may have different values of a variable. Examples of variables are income last year, number of siblings, whether employed, and gender. The values the variable can take form the ***measurement scale***. For gender, for instance, the measurement scale consists of the two labels, female and male. For number of siblings it is 0, 1, 2, 3, ...

The valid statistical methods for a variable depend on its measurement scale. We treat a numerical-valued variable such as annual income differently than a variable with a measurement scale consisting of categories, such as (yes, no) for whether employed. We next present ways to classify variables. The first type refers to whether the measurement scale consists of categories or numbers. Another type refers to the number of levels in that scale.

Quantitative and Categorical Variables

A variable is called ***quantitative*** when the measurement scale has numerical values. The values represent different magnitudes of the variable. Examples of quantitative variables are a subject's annual income, number of siblings, age, and number of years of education completed.

A variable is called ***categorical*** when the measurement scale is a set of categories. For example, marital status, with categories (single, married, divorced, widowed), is categorical. For Canadians, the province of residence is categorical, with the categories Alberta, British Columbia, and so on. Other categorical variables are whether employed (yes, no), primary clothes shopping destination (local mall, local downtown, Internet, other), favorite type of music (classical, country, folk, jazz, rock), religious affiliation (Protestant, Catholic, Jewish, Muslim, other, none), and political party preference.

For categorical variables, distinct categories differ in quality, not in numerical magnitude. Categorical variables are often called ***qualitative***. We distinguish between categorical and quantitative variables because different statistical methods apply to each type. Some methods apply to categorical variables and others apply to quantitative variables. For example, the *average* is a statistical summary for a quantitative variable, because it uses numerical values. It's possible to find the average for a quantitative variable such as income, but not for a categorical variable such as religious affiliation or favorite type of music.

Nominal, Ordinal, and Interval Scales of Measurement

For a quantitative variable, the possible numerical values are said to form an ***interval*** scale. Interval scales have a specific numerical distance or *interval* between each pair of levels. Annual income is usually measured on an interval scale. The interval between \$40,000 and \$30,000, for instance, equals \$10,000. We can compare outcomes in terms of how much larger or how much smaller one is than the other.

Categorical variables have two types of scales. For the categorical variables mentioned in the previous subsection, the categories are unordered. The scale does not have a "high" or "low" end. The categories are then said to form a ***nominal scale***. For another example, a variable measuring primary mode of transportation to work might use the nominal scale with categories (automobile, bus, subway, bicycle, walk).

Although the different categories are often called the *levels* of the scale, for a nominal variable no level is greater than or smaller than any other level. Names or labels such as "automobile" and "bus" for mode of transportation identify the categories but do not represent different magnitudes. By contrast, each possible value of a quantitative variable is *greater than* or *less than* any other possible value.

A third type of scale falls, in a sense, between nominal and interval. It consists of categorical scales having a natural *ordering* of values. The levels form an ***ordinal scale***. Examples are social class (upper, middle, lower), political philosophy (very liberal, slightly liberal, moderate, slightly conservative, very conservative), government spending on the environment (too little, about right, too much), and frequency of religious activity (never, less than once a month, about 1–3 times a month, every week, more than once a week). These scales are not nominal, because the categories are ordered. They are not interval, because there is no defined distance between levels. For example, a person categorized as very conservative is *more* conservative than a person categorized as slightly conservative, but there is no numerical value for *how much more* conservative that person is.

In summary, for ordinal variables the categories have a natural ordering, whereas for nominal variables the categories are unordered. The scales refer to the actual measurement and not to the phenomena themselves. *Place of residence* may indicate a geographic place name such as a county (nominal), the distance of that place from a point on the globe (interval), the size of the place (interval or ordinal), or other kinds of variables.

Quantitative Aspects of Ordinal Data

As we've discussed, levels of nominal scales are qualitative, varying in quality, not in quantity. Levels of interval scales are quantitative, varying in magnitude. The position of ordinal scales on the quantitative–qualitative classification is fuzzy. Because their scale is a set of categories, they are often analyzed using the same methods as nominal scales. But in many respects, ordinal scales more closely resemble interval scales. They possess an important quantitative feature: Each level has a *greater* or *smaller* magnitude than another level.

Some statistical methods apply specifically to ordinal variables. Often, though, it's helpful to analyze ordinal scales by assigning numerical scores to categories. By treating ordinal variables as interval rather than nominal, we can use the more powerful methods available for quantitative variables.

For example, course grades (such as A, B, C, D, E) are ordinal. But we treat them as interval when we assign numbers to the grades (such as 4, 3, 2, 1, 0) to compute a grade point average. Treating ordinal variables as interval requires good judgment in assigning scores. In doing this, you can conduct a "sensitivity analysis" by checking whether conclusions would differ in any significant way for other choices of the scores.

Discrete and Continuous Variables

One other way to classify a variable also helps determine which statistical methods are appropriate for it. This classification refers to the number of values in the measurement scale.

Discrete and Continuous Variables

A variable is ***discrete*** if its possible values form a set of separate numbers, such as 0, 1, 2, 3, ... It is ***continuous*** if it can take an infinite continuum of possible real number values.

Examples of discrete variables are the number of siblings and the number of visits to a physician last year. Any variable phrased as "the number of ..." is discrete, because it is possible to list its possible values $\{0, 1, 2, 3, 4, \ldots\}$.

Examples of continuous variables are height, weight, and the amount of time it takes to read a passage of a book. It is impossible to write down all the distinct potential values, since they form an interval of infinitely many values. The amount of time needed to read a book, for example, could take on the value 8.6294473... hours.

Discrete variables have a basic unit of measurement that cannot be subdivided. For example, 2 and 3 are possible values for the number of siblings, but 2.5716 is not. For a continuous variable, by contrast, between any two possible values there is always another possible value. For example, age is continuous in the sense that an individual does not age in discrete jumps. At some well-defined point during the year in which you age from 21 to 22, you are 21.3851 years old, and similarly for every other real number between 21 and 22. A continuous, infinite collection of age values occurs between 21 and 22 alone.

Any variable with a finite number of possible values is discrete. All categorical variables, nominal or ordinal, are discrete, having a finite set of categories. Quantitative variables can be discrete or continuous; age is continuous, and number of siblings is discrete.

For quantitative variables the distinction between discrete and continuous variables can be blurry, because of how variables are actually measured. In practice, we round continuous variables when measuring them, so the measurement is actually discrete. We say that an individual is 21 years old whenever that person's age is somewhere between 21 and 22. On the other hand, some variables, although discrete, have a very large number of possible values. In measuring annual family income in dollars, the potential values are 0, 1, 2, 3, ..., up to some very large value in many millions.

What's the implication of this? Statistical methods for discrete variables are mainly used for quantitative variables that take relatively few values, such as the number of times a person has been married. Statistical methods for continuous variables are used for quantitative variables that can take lots of values, regardless of whether they are theoretically continuous or discrete. For example, statisticians treat variables such as age, income, and IQ as continuous.

In summary,

- Variables are either *quantitative* (numerical valued) or *categorical*. Quantitative variables are measured on an *interval* scale. Categorical variables with unordered categories have a *nominal* scale, and categorical variables with ordered categories have an *ordinal* scale.
- Categorical variables (nominal or ordinal) are *discrete*. Quantitative variables can be either discrete or continuous. In practice, quantitative variables that can take lots of values are treated as *continuous*.

Figure 2.1 summarizes the types of variables, in terms of the (quantitative, categorical), (nominal, ordinal, interval), and (continuous, discrete) classifications.

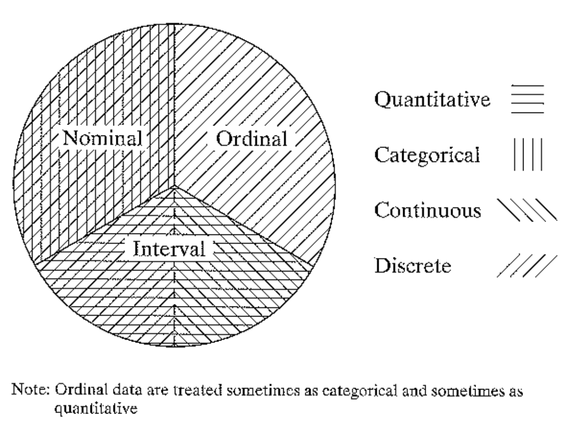

FIGURE 2.1: Summary of Quantitative–Categorical, Nominal–Ordinal–Interval, Continuous–Discrete Classifications

2.2 RANDOMIZATION

Inferential statistical methods use sample statistics to make predictions about population parameters. The quality of the inferences depends on how well the sample represents the population. This section introduces an important sampling method that incorporates ***randomization***, the mechanism for achieving good sample representation.

Simple Random Sampling

Subjects of a population to be sampled could be individuals, families, schools, cities, hospitals, records of reported crimes, and so on. *Simple random sampling* is a method of sampling for which every possible sample has equal chance of selection.

Let n denote the number of subjects in the sample, called the ***sample size***.

> **Simple Random Sample**
>
> A ***simple random sample*** of n subjects from a population is one in which each possible sample of that size has the same probability (chance) of being selected.

For instance, suppose a researcher administers a questionnaire to one randomly selected adult in each of several households. A particular household contains four adults—mother, father, aunt, and uncle—identified as M, F, A, and U. For a simple random sample of $n = 1$ adult, each of the four adults is equally likely to be interviewed. You could select one by placing the four names on four identical ballots and selecting one blindly from a hat. For a simple random sample of $n = 2$ adults, each possible sample of size two is equally likely. The six potential samples are (M, F), (M, A), (M, U), (F, A), (F, U), and (A, U). To select the sample, you blindly select two ballots from the hat.

A simple random sample is often just called a ***random sample***. The *simple* adjective is used to distinguish this type of sampling from more complex sampling schemes presented in Section 2.4 that also have elements of randomization.

Why is it a good idea to use random sampling? Because everyone has the same chance of inclusion in the sample, so it provides fairness. This reduces the chance that the sample is seriously biased in some way, leading to inaccurate inferences about the population. Most inferential statistical methods assume randomization of the sort provided by random sampling.

How to Select a Simple Random Sample

To select a random sample, we need a list of all subjects in the population. This list is called the ***sampling frame***. Suppose you plan to sample students at your school. The population is all students at the school. One possible sampling frame is the student directory.

The most common method for selecting a random sample is to (1) number the subjects in the sampling frame, (2) generate a set of these numbers randomly, and (3) sample the subjects whose numbers were generated. Using *random numbers* to select the sample ensures that each subject has an equal chance of selection.

> **Random Numbers**
>
> ***Random numbers*** are numbers that are computer generated according to a scheme whereby each digit is equally likely to be any of the integers $0, 1, 2, \ldots, 9$ and does not depend on the other digits generated.

TABLE 2.1: Part of a Table of Random Numbers

Line/Col.	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1	10480	15011	01536	02011	81647	91646	69179	14194
2	22368	46573	25595	85393	30995	89198	27982	53402
3	24130	48360	22527	97265	76393	64809	15179	24830
4	42167	93093	06243	61680	07856	16376	39440	53537
5	37570	39975	81837	16656	06121	91782	60468	81305
6	77921	06907	11008	42751	27756	53498	18602	70659

Source: Abridged from William H. Beyer, ed., *Handbook of Tables for Probability and Statistics*, 2nd ed., © The Chemical Rubber Co., 1968. Used by permission of the Chemical Rubber Co.

Table 2.1 shows a table containing random numbers. The numbers fluctuate according to no set pattern. Any particular number has the same chance of being a 0, 1, 2, ..., or 9. The numbers are chosen independently, so any one digit chosen has no influence on any other selection. If the first digit in a row of the table is a 9, for instance, the next digit is still just as likely to be a 9 as a 0 or 1 or any other number. Random numbers are available in published tables and can be generated with software and many statistical calculators.

Suppose you want to select a simple random sample of $n = 100$ students from a university student body of size 30,000. The sampling frame is a directory of these students. Select the students by using five-digit sequences to identify them, as follows:

1. Assign the numbers 00001 to 30000 to the students in the directory, using 00001 for the first student in the list, 00002 for the second student, and so on.
2. Starting at any point in the random number table, or by generating random numbers using software or a calculator, choose successive five-digit numbers until you obtain 100 distinct numbers between 00001 and 30000.
3. Include in the sample the students with assigned numbers equal to the random numbers selected.

For example, for the first column of five-digit numbers in Table 2.1, the first three random numbers are 10480, 22368, and 24130. The first three students selected are those numbered 10480, 22368, and 24130.

In selecting the 100 five-digit numbers, skip numbers greater than 30000, such as the next three five-digit numbers in Table 2.1, since no student in the directory has an assigned number that large. After using the first column of five-digit numbers, move to the next column of numbers and continue. If the population size were between 1000 and 9999, you would use four digits at a time. The column (or row) from which you begin selecting the numbers does not matter, since the numbers have no set pattern. Most statistical software can do this all for you.

Collecting Data with Sample Surveys

Many studies select a sample of people from a population and interview them to collect data. This method of data collection is called a ***sample survey***. The interview could be a personal interview, telephone interview, or self-administered questionnaire.

The General Social Survey (GSS) is an example of a sample survey. It gathers information using personal interviews of a random sample of subjects from the U.S. adult population to provide a snapshot of that population. (The survey does not use *simple* random sampling but rather a method discussed later in the chapter that

incorporates multiple stages and clustering but is designed to give each family the same chance of inclusion.) National polls such as the Gallup Poll are also sample surveys. They usually use telephone interviews. Since it is often difficult to obtain a sampling frame, many telephone interviews obtain the sample with *random digit dialing*.

A variety of problems can cause responses from a sample survey to tend to favor some parts of the population over others. Then results from the sample are not representative of the population. We'll discuss some potential problems in Section 2.3.

Collecting Data with an Experiment

In some studies, data result from a planned *experiment*. The purpose of most experiments is to compare responses of subjects on some outcome measure, under different conditions. Those conditions are levels of a variable that can influence the outcome. The scientist has the experimental control of being able to assign subjects to the conditions.

For instance, the conditions might be different drugs for treating some illness. The conditions are called ***treatments***. To conduct the experiment, the researcher needs a plan for assigning subjects to the treatments. These plans are called ***experimental designs***. Good experimental designs use randomization to determine which treatment a subject receives.

In the late 1980s, the Physicians' Health Study Research Group at Harvard Medical School designed an experiment to analyze whether regular intake of aspirin reduces mortality from heart disease. Of about 22,000 male physicians, half were randomly chosen to take an aspirin every other day. The remaining half took a placebo, which had no active agent. After five years, rates of heart attack were compared. By using randomization to determine who received which treatment, the researchers knew the groups would roughly balance on factors that could affect heart attack rates, such as age and quality of health. If the physicians could decide on their own which treatment to take, the groups might have been out of balance on some important factor. Suppose, for instance, younger physicians were more likely to select aspirin. Then, a lower heart attack rate among the aspirin group could occur merely because younger subjects are less likely to suffer heart attacks.

Collecting Data with an Observational Study

In social research, it is rarely possible to conduct experiments. It's not possible to randomly assign subjects to the groups we want to compare, such as levels of gender or race or educational level or annual income. Many studies merely *observe* the outcomes for available subjects on the variables without any experimental manipulation of the subjects. Such studies are called ***observational studies***. The researcher measures subjects' responses on the variables of interest but has no experimental control over the subjects.

With observational studies, comparing groups is difficult because the groups may be imbalanced on variables that affect the outcome. This is true even with random sampling. For instance, suppose we plan to compare black students, Hispanic students, and white students on some standardized test. If white students have a higher average score, a variety of variables might account for that difference, such as parents' education or parents' income or quality of school attended. This makes it difficult to compare groups with observational studies, especially when some key variables may not have been measured in the study.

Establishing cause and effect is central to science. But it's not possible to establish cause and effect definitively with a nonexperimental study, whether it be an observational study with an available sample or a sample survey using random sampling. With an observational study, there's the strong possibility that the sample does not well reflect the population. With an observational study or a sample survey, there's always the possibility that some unmeasured variable could be responsible for patterns observed in the data. With an experiment that randomly assigns subjects to treatments, those treatments should roughly balance on any unmeasured variables. For example, in the heart attack study mentioned above, the doctors taking aspirin would not tend to be younger or of better health than the doctors taking placebo. Because a randomized experiment balances the groups being compared on other factors, it's possible to study cause and effect more accurately with an experiment than with an observational study.

Whether or not a study is experimental, it's important to incorporate randomization in any study that plans to make inferences. This randomization could take the form of randomly selecting a sample for a survey, or randomly allocating subjects to different treatments for an experimental study.

2.3 SAMPLING VARIABILITY AND POTENTIAL BIAS

Even if a study wisely uses randomization, the results of the study still depend on which subjects are sampled. Two researchers who separately select random samples from some population may have little overlap, if any, between the two sample memberships. Therefore, the values of sample statistics will differ for the two samples, and the results of analyses based on these samples may differ.

Sampling Error

Suppose the Gallup, Harris, Zogby, and Pew polling organizations each randomly sample 1000 adult Canadians, in order to estimate the percentage of Canadians who give the prime minister's performance in office a favorable rating. Based on the samples they select, perhaps Gallup reports an approval rating of 63%, Harris reports 68%, Zogby 65%, and Pew 64%. These differences could reflect slightly different question wording. But even if the questions are worded exactly the same, the percentages would probably differ somewhat because the samples are different.

For conclusions based on statistical inference to be worthwhile, we should know the potential *sampling error*—how much the statistic differs from the parameter it predicts because of the way results naturally exhibit variation from sample to sample.

> **Sampling Error**
>
> The ***sampling error*** of a statistic equals the error that occurs when we use a statistic based on a sample to predict the value of a population parameter.

Suppose that the actual percentage of the population of adult Canadians who give the prime minister a favorable rating is 66%. Then the Gallup organization, which predicted 63%, had a sampling error of $63\% - 66\% = -3\%$. The Harris organization, which predicted 68%, had a sampling error of $68\% - 66\% = 2\%$. In practice, the sampling error is unknown, because the values of population parameters are unknown.

Random sampling protects against bias, in the sense that the sampling error tends to fluctuate about 0, sometimes being positive (as in the Harris poll) and sometimes being negative (as in the Gallup poll). Random sampling also allows us to predict the likely size of the sampling error. For sample sizes of about 1000, we'll see that

the sampling error for estimating percentages is usually no greater than plus or minus 3%. This bound is the *margin of error*. Variability also occurs in the values of sample statistics with nonrandom sampling, but the extent of the sampling error is not predictable as it is with random samples.

Sampling Bias: Nonprobability Sampling

Other factors besides sampling error can cause results to vary from sample to sample. These factors can also possibly cause bias. We next discuss three types of bias. The first is called ***sampling bias***.

For simple random sampling, each possible sample of n subjects has the same probability of selection. This is a type of ***probability sampling*** method, meaning that the probability any particular sample will be selected is known. Inferential statistical methods assume probability sampling. ***Nonprobability sampling*** methods are ones for which it is not possible to determine the probabilities of the possible samples. Inferences using such samples have unknown reliability and result in ***sampling bias***.

The most common nonprobability sampling method is ***volunteer sampling***. As the name implies, subjects volunteer to be in the sample. But the sample may poorly represent the population and yield misleading conclusions. For instance, a mail-in questionnaire published in *TV Guide* posed the question, "Should the President have the Line Item Veto to eliminate waste?" Of those who responded, 97% said yes. For the same question posed to a random sample, 71% said yes.[1]

Examples of volunteer sampling are visible any day on many Internet sites and television news programs. Viewers register their opinions on an issue by voting over the Internet. The viewers who respond are unlikely to be a representative cross section, but will be those who can easily access the Internet and who feel strongly enough to respond. Individuals having a particular opinion might be much more likely to respond than individuals having a different opinion. For example, one night the ABC program *Nightline* asked viewers whether the United Nations should continue to be located in the United States. Of more than 186,000 respondents, 67% wanted the United Nations out of the United States. At the same time, a poll using a random sample of about 500 respondents estimated the population percentage to be about 28%. Even though the random sample had a much smaller size, it is far more trustworthy.

A large sample does not help with volunteer sampling—the bias remains. In 1936, the newsweekly *Literary Digest* sent over 10 million questionnaires in the mail to predict the outcome of the presidential election. The questionnaires went to a relatively wealthy segment of society (those having autos or telephones), and fewer than 25% were returned. The journal used these to predict an overwhelming victory by Alfred Landon over Franklin Roosevelt. The opposite result was predicted by George Gallup with a much smaller sample in the first scientific poll taken for this purpose. In fact, Roosevelt won in a landslide.

Unfortunately, volunteer sampling is sometimes necessary. This is often true in medical studies. Suppose a study plans to investigate how well a new drug performs compared to a standard drug, for subjects who suffer from high blood pressure. The researchers are not going to be able to find a sampling frame of all who suffer from high blood pressure and take a simple random sample of them. They may, however, be able to sample such subjects at certain medical centers or using volunteers. Even then, randomization should be used wherever possible. For the study patients, they can randomly select who receives the new drug and who receives the standard one.

[1] D. M. Wilbur, *Public Perspective*, available at roperweb.ropercenter.uconn.edu, May–June 1993.

Even with random sampling, sampling bias can occur. One case is when the sampling frame suffers from ***undercoverage***: It lacks representation from some groups in the population. A telephone survey will not reach prison inmates or homeless people or people too poor to afford a telephone, whereas families that have many phones will tend to be over-represented. Responses by those not having a telephone might tend to be quite different from those actually sampled, leading to biased results.

Response Bias

In a survey, the way a question is worded or asked can have a large impact on the results. For example, when a *New York Times*/CBS News poll in 2006 asked whether the interviewee would be in favor of a new gasoline tax, only 12% said yes. When the tax was presented as reducing U.S. dependence on foreign oil, 55% said yes, and when asked about a gas tax that would help reduce global warming, 59% said yes.[2]

Poorly worded or confusing questions result in ***response bias***. Even the order in which questions are asked can influence the results dramatically. During the Cold War, a study asked, "Do you think the U.S. should let Russian newspaper reporters come here and send back whatever they want?" and "Do you think Russia should let American newspaper reporters come in and send back whatever they want?" The percentage of yes responses to the first question was 36% when it was asked first and 73% when it was asked second.[3]

In an interview, characteristics of the interviewer may result in response bias. Respondents might lie if they think their belief is socially unacceptable. They may be more likely to give the answer that they think the interviewer prefers. An example is provided by a study on the effect of the interviewer's race. Following a phone interview, respondents were asked whether they thought the interviewer was black or white (all were actually black). Perceiving a white interviewer resulted in more conservative opinions. For example, 14% agreed that "American society is fair to everyone" when they thought the interviewer was black, but 31% agreed to this when they thought the interviewer was white.[4]

Nonresponse Bias: Missing Data

Some subjects who are supposed to be in the sample may refuse to participate, or it may not be possible to reach them. This results in the problem of ***nonresponse bias***. If only half the intended sample was actually observed, we should worry about whether the half not observed differ from those observed in a way that causes biased results. Even if we select the sample randomly, the results are questionnable if there is substantial nonresponse, say, over 20%.

For her book *Women in Love*, author Shere Hite surveyed women in the United States. One of her conclusions was that 70% of women who had been married at least five years have extramarital affairs. She based this conclusion on responses to questionnaires returned by 4500 women. This sounds like an impressively large sample. However, the questionnaire was mailed to about 100,000 women. We cannot know whether the 4.5% of the women who responded were representative of the 100,000 who received the questionnaire, much less the entire population of American women. This makes it dangerous to make an inference to the larger population.

[2] Column by T. Friedman, *New York Times*, March 2, 2006.
[3] See Crosson (1994).
[4] *Washington Post*, June 26, 1995.

Missing data is a problem in almost all large studies. Some subjects do not provide responses for some of the variables measured. Even in censuses, which are designed to observe everyone in a country, some people are not observed or fail to cooperate. Most software ignores cases for which observations are missing for at least one variable used in an analysis. This results in wasted information and possible bias. Statisticians have recently developed methods that replace missing observations by predicted values based on patterns in the data. See Allison (2002) for an introduction to ways of dealing with missing data.

Summary of Types of Bias

In summary, sample surveys have potential sources of bias:

- ***Sampling bias*** occurs from using nonprobability samples or having undercoverage.
- ***Response bias*** occurs when the subject gives an incorrect response (perhaps lying), or the question wording or the way the interviewer asks the questions is confusing or misleading.
- ***Nonresponse bias*** occurs when some sampled subjects cannot be reached or refuse to participate or fail to answer some questions.

In any study, carefully assess the scope of conclusions. Evaluate critically the conclusions by noting the makeup of the sample. How was the sample selected? How large was it? How were the questions worded? Who sponsored and conducted the research? The less information that is available, the less you should trust it.

Finally, be wary of any study that makes inferences to a broader population than is justified by the sample chosen. Suppose a psychologist performs an experiment using a random sample of students from an introductory psychology course. With statistical inference, the sample results generalize to the population of all students in the class. For the results to be of wider interest, the psychologist might claim that the conclusions extend to *all* college students, to all young adults, or even to all adults. These generalizations may well be wrong, because the sample may differ from those populations in fundamental ways, such as in average age or socioeconomic status.

2.4 OTHER PROBABILITY SAMPLING METHODS*

Section 2.2 introduced ***simple random sampling*** and explained its importance to statistical inference. In practice, other probability sampling methods that have elements of randomness are sometimes preferable to simple random sampling or are simpler to obtain.

Systematic Random Sampling

Systematic random sampling selects a subject near the beginning of the sampling frame list, skips several names and selects another subject, skips several more names and selects the next subject, and so forth. The number of names skipped at each stage depends on the desired sample size. Here's how it's done:

Systematic Random Sample

Denote the sample size by n and the population size by N. Let $k = N/n$, the population size divided by the sample size. A ***systematic random sample*** (1) selects a subject at random from the first k names in the sampling frame, and (2) selects every kth subject listed after that one. The number k is called the *skip number*.

Suppose you want a systematic random sample of 100 students from a population of 30,000 students listed in a campus directory. Then, $n = 100$ and $N = 30{,}000$, so $k = 30{,}000/100 = 300$. The population size is 300 times the sample size, so you need to select one of every 300 students. You select one student at random, using random numbers, from the first 300 students in the directory. Then you select every 300th student after the one selected randomly. This produces a sample of size 100. The first three digits in Table 2.1 are 104, which falls between 001 and 300, so you first select the student numbered 104. The numbers of the other students selected are $104 + 300 = 404$, $404 + 300 = 704$, $704 + 300 = 1004$, $1004 + 300 = 1304$, and so on. The 100th student selected is listed in the last 300 names in the directory.

In sampling from a sampling frame, it's simpler to select a systematic random sample than a simple random sample because it uses only one random number. This method typically provides as good a representation of the population, because for alphabetic listings such as directories of names, values of most variables fluctuate randomly through the list. With this method, statistical formulas based on simple random sampling are usually valid.

A systematic random sample is not a simple random sample, because all samples of size n are not equally likely. For instance, unlike in a simple random sample, two subjects listed next to each other on the list cannot both appear in the sample.

Stratified Random Sampling

Another probability sampling method, useful in social science research for studies comparing groups, is *stratified sampling*.

Stratified Random Sample

A ***stratified random sample*** divides the population into separate groups, called ***strata***, and then selects a simple random sample from each stratum.

Suppose a study in Cambridge, Massachusetts plans to compare the opinions of registered Democrats and registered Republicans about whether government should guarantee health care to all citizens. Stratifying according to political party registration, the study selects a random sample of Democrats and another random sample of Republicans.

Stratified random sampling is called ***proportional*** if the sampled strata proportions are the same as those in the entire population. For example, if 90% of the population of interest is Democrat and 10% is Republican, then the sampling is proportional if the sample size for Democrats is nine times the sample size for Republicans.

Stratified random sampling is called ***disproportional*** if the sampled strata proportions differ from the population proportions. This is useful when the population size for a stratum is relatively small. A group that comprises a small part of the population may not have enough representation in a simple random sample to allow precise inferences. It is not possible to compare accurately Democrats to Republicans, for example, if only 10 people in a sample size of 100 are Republican. By contrast, a disproportional stratified sample size of 100 might randomly sample 50 Democrats and 50 Republicans.

To implement stratification, we must know the stratum into which each subject in the sampling frame belongs. This usually restricts the variables that can be used for forming the strata. The variables must have strata that are easily identifiable. For example, it would be easy to select a stratified sample of a school population using grade level as the stratification variable, but it would be difficult to prepare an adequate sampling frame of city households stratified by household income.

Cluster Sampling

Simple, systematic, and stratified random sampling are often difficult to implement, because they require a complete sampling frame. Such lists are easy to obtain for sampling cities or hospitals or schools, for example, but more difficult for sampling individuals or families. *Cluster sampling* is useful when a complete listing of the population is not available.

Cluster Random Sample

Divide the population into a large number of ***clusters***, such as city blocks. Select a simple random sample of the clusters. Use the subjects in those clusters as the sample.

For example, a study might plan to sample about 1% of the families in a city, using city blocks as clusters. Using a map to identify city blocks, it could select a simple random sample of 1% of the blocks and then sample every family on each block. A study of patient care in mental hospitals in Ontario could first randomly sample mental hospitals (the clusters) and then collect data for patients within those hospitals.

What's the difference between a stratified sample and a cluster sample? A stratified sample uses *every* stratum. The strata are usually groups we want to compare. By contrast, a cluster sample uses a *sample* of the clusters, rather than all of them. In cluster sampling, clusters are merely ways of easily identifying groups of subjects. The goal is not to compare the clusters but rather to use them to obtain a sample. Most clusters are not represented in the eventual sample.

Figure 2.2 illustrates the distinction among sampling subjects (simple random sample), sampling clusters of subjects (cluster random sample), and sampling subjects from within strata (stratified random sample). The figure depicts ways to survey 40 students at a school, to make comparisons among Freshmen, Sophomores, Juniors, and Seniors.

Multistage Sampling

When conducting a survey for predicting elections, the Gallup Organization often identifies election districts as clusters and takes a simple random sample of them. But then it also takes a simple random sample of households within each selected election

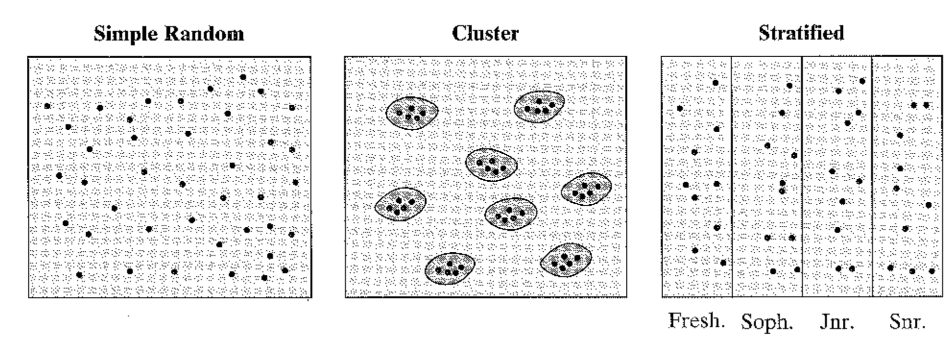

FIGURE 2.2: Ways of Randomly Sampling 40 Students. The figure is a schematic for a simple random sample, a cluster random sample of 8 clusters of students who live together, and a stratified random sample of 10 students from each class (Fr, So, Ju, Sr).

district. This is more feasible than sampling *every* household in the chosen districts. This is an example of *multistage sampling*, which uses combinations of sampling methods.

Here's an example of a multistage sample:

- Treat counties (or census tracts) as clusters and select a random sample of a certain number of them.
- Within each county selected, take a cluster random sample of square-block regions.
- Within each region selected, take a systematic random sample of every tenth house.
- Within each house selected, select one adult at random for the sample.

Multistage samples are common in social science research. They are simpler to implement than simple random sampling but provide a broader sampling of the population than a single method such as cluster sampling.

For statistical inference, stratified samples, cluster samples, and multistage samples use different formulas from the ones in this book. Cluster sampling requires a larger sample to achieve as much inferential precision as simple random sampling. Observations within clusters tend to be similar, because of the tendency of subjects living near one another to have similar values on opinion issues and on economic and demographic variables such as age, income, race, and occupation. So we need more data to obtain a representative cross section. By contrast, the results for stratified sampling may be more precise than those stated in this textbook for simple random sampling. Books specializing in sampling methodology provide further details (Scheaffer et al., 2005; Thompson, 2002).

2.5 CHAPTER SUMMARY

Statistical methods analyze data on ***variables***, which are characteristics that vary among subjects. The statistical methods used depend on the type of variable:

- Numerically measured variables, such as family income and number of children in a family, are ***quantitative***. They are measured on an *interval scale*.
- Variables taking value in a set of categories are ***categorical***. Those measured with unordered categories, such as religious affiliation and province of residence, have a *nominal scale*. Those measured with ordered categories, such as social class and political ideology, have an *ordinal scale* of measurement.
- Variables are also classified as ***discrete***, having possible values that are a set of separate numbers (such as 0, 1, 2, ...), or ***continuous***, having a continuous, infinite set of possible values. Categorical variables, whether nominal or ordinal, are discrete. Quantitative variables can be of either type, but in practice are treated as continuous if they can take a large number of values.

Much social science research uses ***observational studies***, which use available subjects to observe variables of interest. One should be cautious in attempting to conduct inferential analyses with data from such studies. Inferential statistical methods require ***probability samples***, which incorporate randomization in some way. Random sampling allows control over the amount of ***sampling error***, which describes how results can vary from sample to sample. Random samples are much more likely to be representative of the population than are nonprobability samples such as volunteer samples.

- For a ***simple random sample***, every possible sample of size n has the same chance of selection.

- Here are other examples of probability sampling: ***Systematic*** random sampling takes every kth subject in the sampling frame list. ***Stratified*** random sampling divides the population into groups (strata) and takes a random sample from each stratum. ***Cluster*** random sampling takes a random sample of clusters of subjects (such as city blocks) and uses subjects in those clusters as the sample. ***Multistage*** sampling uses combinations of these methods.

Chapter 3 introduces statistics for describing samples and corresponding parameters for describing populations. Hence, its focus is on *descriptive statistics*.

PROBLEMS

Practicing the Basics

2.1. Explain the difference between
- **(a)** Discrete and continuous variables
- **(b)** Categorical and quantitative variables
- **(c)** Nominal and ordinal variables

Why do these distinctions matter for statistical analysis?

2.2. Identify each variable as categorical or quantitative:
- **(a)** Number of pets in family
- **(b)** County of residence
- **(c)** Choice of auto (domestic or import)
- **(d)** Distance (in miles) commute to work
- **(e)** Choice of diet (vegetarian, nonvegetarian)
- **(f)** Time spent in previous month browsing the World Wide Web
- **(g)** Ownership of personal computer (yes, no)
- **(h)** Number of people you have known with AIDS (0, 1, 2, 3, 4 or more)
- **(i)** Marriage form of a society (monogamy, polygyny, polyandry)

2.3. Which scale of measurement (nominal, ordinal, or interval) is most appropriate for
- **(a)** Attitude toward legalization of marijuana (favor, neutral, oppose)
- **(b)** Gender (male, female)
- **(c)** Number of children in family (0, 1, 2, ...)
- **(d)** Political party affiliation (Democrat, Republican, Independent)
- **(e)** Religious affiliation (Catholic, Jewish, Protestant, Muslim, other)
- **(f)** Political philosophy (very liberal, somewhat liberal, moderate, somewhat conservative, very conservative)
- **(g)** Years of school completed (0, 1, 2, 3, ...)
- **(h)** Highest degree attained (none, high school, bachelor's, master's, doctorate)
- **(i)** College major (education, anthropology, physics, sociology, ...)
- **(j)** Test score (0–100 range for scores)
- **(k)** Employment status (employed full time, employed part time, unemployed)

2.4. Which scale of measurement is most appropriate for
- **(a)** Occupation (plumber, teacher, secretary, ...)
- **(b)** Occupational status (blue collar, white collar)
- **(c)** Social status (lower, middle, upper class)
- **(d)** Statewide murder rate (number of murders per 1000 population)
- **(e)** County population size (number of people)
- **(f)** Population growth rate (in percentages)
- **(g)** Community size (rural, small town, large town, small city, large city)
- **(h)** Annual income (thousands of dollars per year)
- **(i)** Attitude toward affirmative action (favorable, neutral, unfavorable)
- **(j)** Lifetime number of sexual partners

2.5. Which scale of measurement is most appropriate for "attained education" measured as
- **(a)** Number of years (0, 1, 2, 3, ...)
- **(b)** Grade level (elementary school, middle school, high school, college, graduate school)
- **(c)** School type (public school, private school)

2.6. Give an example of a variable that is **(a)** categorical, **(b)** quantitative, **(c)** ordinal scale, **(d)** nominal scale, **(e)** discrete, **(f)** continuous, **(g)** quantitative and discrete.

2.7. A poll conducted by YouGov for the British newspaper *The Daily Telegraph* in June 2006 asked a random sample of 1962 British adults several questions about their image of the U.S. One question asked, "How would you rate George W. Bush as a world leader?" The possible choices were (great leader, reasonably satisfactory leader, pretty poor leader, terrible leader).
- **(a)** Is this four-category variable nominal, or ordinal? Why?
- **(b)** Is this variable continuous, or discrete? Why?
- **(c)** Of the 93% of the sample who responded, the percentages in the four categories were 1% (great leader), 16%, 37%, 46% (terrible leader). Are these values statistics, or parameters? Why?

2.8. A survey asks subjects to rate five issues according to their importance in determining voting intention for U.S. senator, using the scale (very important, somewhat important, unimportant). The issues are foreign policy, unemployment, inflation, the arms race, and civil rights. The evaluations can be treated as five variables: foreign policy evaluation, unemployment evaluation, and so on. These variables represent what scale of measurement?

2.9. Which of the following variables could theoretically be measured on a continuous scale? **(a)** Method of contraception used, **(b)** length of time of residence in a state, **(c)** task completion time, **(d)** intelligence, **(e)** authoritarianism, **(f)** alienation, **(g)** county of residence.

2.10. Which of the following variables are continuous when the measurements are as fine as possible? **(a)** Age of mother, **(b)** number of children in family, **(c)** income of spouse, **(d)** population of cities, **(e)** latitude and longitude of cities, **(f)** distance of home from place of employment, **(g)** number of foreign languages spoken.

2.11. A class has 50 students. Use the column of the first two digits in the random number table (Table 2.1) to select a simple random sample of three students. If the students are numbered 01 to 50, what are the numbers of the three students selected?

2.12. A local telephone directory has 400 pages with 130 names per page, a total of 52,000 names. Explain how you could choose a simple random sample of 5 names. Using the second column of Table 2.1 or software or a calculator, select 5 random numbers to identify subjects for the sample.

2.13. Explain whether an experiment or an observational study would be more appropriate to investigate the following:

(a) Whether or not cities with higher unemployment rates tend to have higher crime rates

(b) Whether a Honda Accord or a Toyota Camry gets better gas mileage

(c) Whether or not higher college GPAs tend to occur for students who had higher scores on college entrance exams

(d) Whether or not a special coupon attached to the outside of a catalog makes recipients more likely to order products from a mail-order company

2.14. A study is planned to study whether passive smoking (being exposed to secondhand cigarette smoke on a regular basis) leads to higher rates of lung cancer.

(a) One possible study is to take a sample of children, randomly select half of them for placement in an environment where they are passive smokers, and place the other half in an environment where they are not exposed to smoke. Then 60 years later the observation is whether each has developed lung cancer. Would this study be an experimental study or an observational study? Why?

(b) For many reasons, including time and ethics, it is not possible to conduct the study in (a). Describe a way that *is* possible, and indicate whether it would be an experimental or observational study.

2.15. Table 2.2 shows the result of the 2000 Presidential election and the predictions of several organizations in the days before the election. The sample sizes were typically about 2000. The percentages for each poll do not sum to 100 because of voters reporting as undecided or favoring another candidate.

(a) What factors cause the results to vary somewhat among organizations?

(b) Identify the sampling error for the Gallup poll.

TABLE 2.2

	Predicted Vote		
Poll	Gore	Bush	Nader
Gallup	46	48	4
Harris	47	47	5
ABC	45	48	3
CBS	45	44	4
NBC	44	47	3
Pew Research	47	49	4
Actual vote	48.4	47.9	2.7

Source: www.ncpp.org/

2.16. The BBC in Britain requested viewers to call the network and indicate their favorite poem. Of more than 7500 callers, more than twice as many voted for Rudyard Kipling's *If* than for any other poem. The BBC reported that this was the clear favorite.

(a) Explain what it means to call this a "volunteer sample."

(b) If the BBC truly wanted to determine Brits' favorite poem, how could it more reliably do so?

2.17. A Roper Poll was designed to determine the percentage of Americans who express some doubt that the Nazi Holocaust occurred. In response to the question, "Does it seem possible or does it seem impossible to you that the Nazi extermination of the Jews never happened?" 22% said it was possible the Holocaust never happened. The Roper organization later admitted that the question was worded in a confusing manner. When the poll asked, "Does it seem possible to you that the Nazi extermination of the Jews never happened, or do you feel certain that it happened?" only 1% said it was possible it never happened.[5] Use this example to explain the concept of response bias.

2.18. Refer to Exercise 2.12 about selecting 5 of 52,000 names on 400 pages of a directory.

(a) Select five numbers to identify subjects for a systematic random sample of five names from the directory.

(b) Is cluster sampling applicable? How could it be carried out, and what would be the advantages and disadvantages?

2.19. You plan to sample from the 5000 students at a college, to compare the proportions of men and women who believe that the legal age for alcohol should be changed to 18. Explain how you would proceed if you want a systematic random sample of 100 students.

2.20. You plan to sample from the 3500 undergraduate students enrolled at the University of Rochester, to compare the proportions of female and male students who would like to see the U.S. have a female President.

(a) Suppose that you use random numbers to select students, but you stop selecting females as soon as you have 40, and you stop selecting males as soon as you have 40. Is the resulting sample a simple random sample? Why or why not?

(b) What type of sample is the sample in (a)? What advantage might it have over a simple random sample?

2.21. Clusters versus strata:

(a) With a cluster random sample, do you take a sample of (i) the clusters? (ii) the subjects within every cluster?

(b) With a stratified random sample, do you take a sample of (i) the strata? (ii) the subjects within every stratum?

(c) Summarize the main differences between cluster sampling and stratified sampling in terms of whether you sample the groups or sample from within the groups that form the clusters or strata.

Concepts and Applications

2.22. Refer to the *Student survey* data file introduced in Exercise 1.11 (page 8). For each variable in the data set, indicate whether it is:

(a) Categorical or quantitative

(b) Nominal, ordinal, or interval

2.23. Repeat the previous exercise for the data file created in Exercise 1.12 (page 9).

2.24. You are directing a study to determine the factors that relate to good academic performance at your school.

(a) Describe how you might select a sample of 100 students for the study.

(b) List some variables that you would measure. For each, provide the scale you would use to measure it, and indicate whether statistical analysis could treat it as (i) categorical or quantitative, (ii) nominal, ordinal, or interval, (iii) continuous or discrete.

(c) Give an example of a research question that could be addressed using data on the variables you listed in (b).

2.25. With ***quota sampling*** a researcher stands at a street corner and conducts interviews until obtaining a quota representing the relative sizes of various groups in the population. For instance, the quota might be 50 factory workers, 100 housewives, 60 elderly people, 30 blacks, and so forth. Is this a probability or nonprobability sampling method? Explain, and discuss potential advantages or disadvantages of this method. (Professional pollsters such as Gallup used this method until 1948, when they incorrectly predicted that Dewey would defeat Truman in a landslide in the presidential election.)

2.26. When the Yankelovich polling organization asked,[6] "Should laws be passed to eliminate all possibilities of special interests giving huge sums of money to candidates?" 80% of the sample answered *yes*. When they posed the question, "Should laws be passed to prohibit interest groups from contributing to campaigns, or do groups have a right to contribute to the candidate they support?" only 40% said *yes*. Explain what this example illustrates, and use your answer to differentiate between sampling error and response bias in survey results.

2.27. In each of the following situations, evaluate whether the method of sample selection is appropriate for obtaining information about the population of interest. How would you improve the sample design?

(a) A newspaper wants to determine whether its readers believe that government expenditures should be reduced by cutting benefits for the disabled. They provide an Internet address for readers to vote *yes* or *no*. Based on 1434 Internet votes, they report that 93% of the city's residents believe that benefits should be reduced.

(b) A congresswoman reports that letters to her office are running 3 to 1 in opposition to

[5] *Newsweek*, July 25, 1994.

[6] *Source*: *A Mathematician Reads the Newspaper*, by J. A. Paulos, Basic Books, 1995, p. 15.

the passage of stricter gun control laws. She concludes that approximately 75% of her constituents oppose stricter gun control laws.

(c) An anthropology professor wanted to compare attitudes toward premarital sex of physical science majors and social science majors. She administered a questionnaire to her large class of Anthropology 437, Comparative Human Sexuality. She found no appreciable difference between her physical science and social science majors in their attitudes, so she concluded that the two student groups were about the same in their relative acceptance of premarital sex.

(d) A questionnaire was mailed to a simple random sample of 500 household addresses in a city. Ten were returned as bad addresses, 63 were returned completed, and the rest were not returned. The researcher analyzed the 63 cases and reported that they represent a "simple random sample of city households."

(e) A principal in a large high school is interested in student attitudes toward a proposed achievement test to determine whether a student should graduate. She lists all of the first-period classes, assigning a number to each. Then, using a random number table, she chooses a class at random and interviews every student in that class about the proposed test.

2.28. A content analysis of a daily newspaper studies the percentage of newspaper space devoted to news about entertainment. The sampling frame consists of the daily editions of the newspaper for the previous year. What potential problem might there be in using a systematic sample with skip number equal to 7 or a multiple of 7?

2.29. In a systematic random sample, every subject has the same chance of selection, but the sample is not a simple random sample. Explain why.

2.30. With a total sample of size 100, we want to compare Native Americans to other Americans on the percentage favoring legalized gambling. Why might it be useful to take a disproportional stratified random sample?

2.31. In a cluster random sample with equal-sized clusters, every subject has the same chance of selection. However, the sample is not a simple random sample. Explain why not.

2.32. Find an example of results of an Internet poll. Do you trust the results of the poll? If not, explain why not.

2.33. To sample residents of registered nursing homes in Yorkshire, UK, I construct a list of all nursing homes in the county, which I number from 1 to 110. Beginning randomly, I choose every tenth home on the list, ending up with 11 homes. I then obtain lists of residents from those 11 homes, and I select a simple random sample from each list. What kinds of sampling have I used?

For multiple-choice questions 2.34–2.37, select the best response.

2.34. A simple random sample of size n is one in which:

(a) Every nth member is selected from the population.

(b) Each possible sample of size n has the same chance of being selected.

(c) There must be exactly the same proportion of women in the sample as is in the population.

(d) You keep sampling until you have a fixed number of people having various characteristics (e.g., males, females).

(e) A particular minority group member of the population is less likely to be chosen than a particular majority group member.

(f) All of the above

(g) None of the above

2.35. If we use random numbers to take a simple random sample of 50 students from the 20,000 students at a university,

(a) It is impossible to get the random number 11111, because it is not a random sequence.

(b) If we get 20001 for the first random number, for the second random number that number is less likely to occur than the other possible five-digit random numbers.

(c) The draw 12345 is no more or less likely than the draw 11111.

(d) Since the sample is random, it is *impossible* that it will be non-representative, such as having only females in the sample.

2.36. Crosson (1994, p. 168) described an analysis of published medical studies involving treatments for heart attacks. In the studies having randomization and strong controls for bias, the new therapy provided improved treatment 9% of the time. In studies without randomization or other controls for bias, the new therapy provided improved treatment 58% of the time. Select the correct response(s).

(a) This result suggests it is better not to use randomization in medical studies, because it is harder to show that new ideas are beneficial.

(b) Many newspaper articles that suggest that a particular food, drug, or environmental agent is harmful or beneficial should be viewed skeptically, unless we learn more about the statistical design and analysis for the study.

(c) This result suggests that you should be skeptical about published results of medical studies that are not randomized, controlled studies.

(d) Controlling for biases, both suspected and unsuspected, is necessary in medical research but not in social research, because the social sciences deal in subjective rather than objective truth.

2.37. A recent GSS asked subjects if they supported legalizing abortion in each of seven different circumstances. The percentage who supported legalization varied between 45% (if the woman wants it for any reason) to 92% (if the woman's health is seriously endangered by the pregnancy). This indicates that

(a) Responses can depend greatly on the question wording.

(b) Surveys sample only a small part of the population and can never be trusted.

(c) The sample must not have been randomly selected.

(d) The sample must have had problems with bias resulting from subjects not telling the truth.

2.38. An interviewer stands at an entrance to a popular shopping mall and conducts interviews. True or false: Because we cannot predict who will be interviewed, the sample obtained is an example of a random sample. Explain.

2.39. In a recent Miss America beauty pageant, television viewers could cast their vote on whether to cancel the swimwear parade by phoning a number the network provided. About 1 million viewers called and registered their opinion, of whom 79% said they wanted to see the contestants dressed as bathing beauties. True or false: Since everyone had a chance to call, this was a simple random sample of all the viewers of this program. Explain.

2.40.** An interval scale for which ratios are valid is called a ***ratio scale. Such scales have a well-defined 0 point, so, for instance, one can regard the value 20 as twice the quantity of the value 10. Explain why annual income is measured on a ratio scale, but temperature (in Fahrenheit or Centigrade) is not. Is IQ, as a measure of intelligence, a ratio-scale variable?

*Exercises marked with an asterisk are of greater difficulty or introduce new and optional material.