CHAPTER 1

Introduction

1.1 Introduction

Sample surveys are conducted throughout the world on topics related to agriculture, demography, economics, education, health and welfare, politics, and several other topics of interest. National and international organizations and private agencies collect information through sample surveys on items such as agricultural and industrial productions, employment rates, educational levels, and medical services. Information on the public is also collected on demographic items such as age, marital status, and family size and on socioeconomic characteristics such as religion, ethnicity, and occupation.

Information from these surveys is essential, for instance, for agricultural and economic planning, for making improvements in educational and medical programs, and for providing public services. Estimates and predictions of characteristics such as the number of students graduating from colleges and their starting salaries, information on medical services needed for the young and the old, and opinions of the public on political issues are obtained from data collected through surveys.

In the U.S., the Departments of Agriculture, Commerce, Education, Health, and Labor and other government agencies regularly conduct large-scale surveys on the population. *Statistics Canada*, the *National Sample Surveys* in India, and the United Kingdom routinely conduct agricultural, demographic, and economic surveys. Similar surveys are also conducted at regular intervals in several countries in Africa, Asia, Australia, Europe, and Latin America.

At the international level, the Food and Agricultural Organization (FAO), the World Health Organization (WHO), and other agencies of the United Nations conduct surveys throughout the world. The National Household Survey Capability Program of the UN is a cooperative undertaking in Africa and other countries of the world. Collection and dissemination of data related to agricultural productions, economic

activities, and the demographic and socioeconomic aspects of the populations of these countries are some of the major purposes of this program. The World Fertility Survey (WFS) was conducted by the UN between 1972 and 1984 to collect information on the birth rates, family sizes, and other related characteristics from the households in more than 60 countries around the world.

Offices of the presidents, prime ministers, cabinet members, and legislatures employ sample surveys for ascertaining public opinion on impending legislations and reforms. Political scientists use survey data to study the opinions and attitudes of people on the candidates for public offices, defense budgets, government spending, issues related to war and peace and the like. Exit polls are used for predicting the outcomes of political elections. Opinions on school budgets, library facilities, recreational activities, and other public services in towns and cities are frequently obtained from samples of automobile or voter registrations.

Social scientists analyze information from surveys to examine the living conditions, changes in educational standards and employment rates, and the general welfare of the people. Major newspaper, television, and radio networks throughout the world summarize the information collected from polls and surveys on current topics of interest. During times of earthquakes, epidemics, famines, floods, wars and similar exigencies, sample surveys are usually the only sources for obtaining quick information on the affected public.

Surveys are conducted by market research firms to assess the preferences of consumers regarding the household purchases, public services, and opinions on the quality and prices of consumer products. Samples are routinely used for industrial quality control. Performances and reliabilities of automobiles, televisions, refrigerators, computers, and stereo equipment are examined through selected samples.

Television ratings are based on the information collected from a sample of households. The Dow–Jones Industrial Average is computed from a sample of only 30 from almost 3000 stocks. The S&P 500 index is obtained from a sample of 500 stocks selected from specified groups of the stocks. Similar indices are obtained from samples of particular types of stocks. Sample surveys are also used for auditing financial accounts.

Sampling is frequently employed to estimate the sizes of fish, bird, and animal populations and the effects of environmental pollution and ecological changes on their survival. There are many more uses of sample surveys. Information obtained from properly selected samples and the accompanying statistical estimation and inference procedures can

be vital in the daily activities of the people. The essential aspects of sample surveys are described in the following sections.

1.2 Censuses and surveys

The Decennial Census on the population and housing in the U.S. is conducted by the Bureau of the Census. In this type of census, which is also conducted in several other countries, it is required to count every person in the country and obtain information from all the households on a number of demographic and socioeconomic characteristics. In addition to the population and housing censuses, other types such as the censuses of manufacturing industries and medical services are conducted periodically in many countries. Considerable amounts of money and time are expended for collecting the required information through the censuses and for analyzing it. It was reported that the 1990 census of population on approximately 100 million households in the U.S. was conducted at a cost of more than $2.6 billion and almost a year was needed to assemble the required information.

In the context of sample surveys, a collection or an aggregate of units such as people, households, cities, districts, countries, states, or provinces is called a **finite population**. A sample consists of only a portion of the population units. A census is a 100% sample and it is a **complete count** of the population.

To collect information on the population characteristics, there are distinct advantages in considering a sample instead of a census. First, a sample provides timely information with relatively low cost. Second, a number of errors can occur in a census for collecting, recording, tabulating, and analyzing the vast amount of information. In a carefully conducted survey, such mistakes can be avoided to a large extent. Information on employment rates, educational levels, and other characteristics of the population of national interest is also needed between the census periods. The national and international surveys mentioned in the previous section provide such information. Sample surveys are also conducted during a census to obtain information on variables not included in the census.

Both censuses and sample surveys may fail to include some of the population units, and may not be able to contact some of the included units. Even if they are contacted, it may also become difficult to obtain the required information from some of the contacted units. These types of noncoverage, noncontact, and nonresponse are known as the

nonsampling errors. These errors and errors of observations, measurements, and recording occur in large-scale censuses more frequently than in sample surveys.

1.3 Types of surveys

Many of the national and international surveys are usually of the **multisubject** and **multipurpose** type. For example, a household survey may include a number of demographic and socioeconomic characteristics such as household sizes, incomes of the heads of households, number of children attending elementary schools, and family finances.

In almost all small and large-scale surveys, before selecting the samples, the population is first divided into a fixed number of **strata** consisting, for example, of geographical regions or groups of units defined according to characteristics such as age, education, income level, sex, and race. Samples are selected independently from each stratum. For many types of demographic and economic surveys, the strata consist of states or provinces and cities. For agricultural surveys, strata are usually formed by dividing the entire land into regions or zones. In some cases, each stratum is further divided into substrata; for example, rural and urban areas of a state.

For large-scale surveys, well-defined **clusters**, groups of contiguous units, may be considered in each stratum. Counties in states, districts in provinces, enumeration areas in rural zones are some examples of clusters. Households, individuals, corporations, and farms may form the final sampling units from which the needed information is obtained.

The actual procedure of sampling in some situations may be implemented in two, three, or more stages. In a household survey, at the first stage, a sample of counties may be drawn from a stratum. At the second stage, a sample of blocks or areas may be drawn from each of the selected counties. At the third stage, a sample of dwellings may be drawn from each of the selected blocks. **Multistage sampling** of this type is a common feature of large-scale surveys.

Several surveys conducted by the government and private organizations are repeated periodically. In the **rotation** or successive sampling procedures, observations are made on some of the units selected previously and on some new units sampled at the current period. In **longitudinal** surveys, observations are made on a selected sample, a **panel**, over periods of time.

For any of the above types of surveys, the population units may be selected into the sample with **equal** or **specified** probabilities, **systematically** or through some other suitable procedure. These approaches are presented in detail in the following chapters.

1.4 Sampling frame

The sample is selected from the **frame**, the list of all the units of the population to be surveyed. The frame should contain all the units of the population under consideration, the **target population**. For example, to estimate the educational expenses of students in a college or university, a sample from the frame with only the names of students living in the dormitories and residence halls will be inadequate. It should also include the names of the students living off-campus. The residents of a metropolitan area may be listed in more than one frame; for instance, automobile registrations and city directories. If a population consists of groups or strata, separate lists of the units in each stratum should be available. Similarly, for the multistage procedures, detailed lists for each stage of sampling should be available.

To estimate agricultural production in some countries, the total region under cultivation is divided into **enumeration districts** (EDs) and **enumeration areas** (EAs) and listed in a frame. For the *aerial* surveys conducted through satellite images, the frame is constructed by dividing the map of the area under consideration into square or rectangular grids.

1.5 Questionnaires, interviews, and sample sizes

Questionnaires

In almost every major survey, information on more than one characteristic is solicited from the respondents. The questionnaire should contain all the items in the form of questions to be answered by the respondents or statements to be completed. "How many persons are in your household?" and "How much do you spend each week on (a) food and (b) entertainment?" are examples of the quantitative type of questions. On attitude and opinion type of questions, the answers to be checked by the respondents usually are in the form of *Yes, No*, or *Don't Know*. To record the tastes, likes, and dislikes on the specified characteristics in a survey,

respondents are sometimes asked to express their preferences on a scale ranging from 0 to 5 or 10.

For quite a long time, employment rates were estimated by the U.S. Department of Labor by asking the adult male respondent whether he was employed at the time of the interview, adult female whether she was "keeping house," and the teenage respondent whether he or she was attending school. This method of questioning was found to be underestimating the unemployment rate, especially of women staying home and seeking employment. Starting in January 1994, it was decided to estimate these rates by asking questions with Yes–No answers.

Another illustration that emphasizes the importance of correct words and sentences in a questionnaire is provided by a survey on the daylight saving time in the U.S. Clocks are set back an hour in the fall and advanced an hour in the spring. This practice provides an extra hour of morning sunlight in the fall. In *The New York Times* of October 30, 1993 a reader cited a 1981 survey in which two thirds of the American public favored an extra hour of morning sunlight throughout the year. Another reader responded that the results of such a survey would probably be different if the public were asked whether it would prefer an extra hour of sunlight in the morning or afternoon.

Interviewing methods

Information from the sample units can be obtained through (1) mail, (2) telephone, and (3) face-to-face personal interviews. In several countries, postal and telephone services are extensively used for surveys. Personal interviews are usually the only means of conducting surveys in the rural areas of many countries. Computer aided telephone interviewing (CATI), in which the interviewer equipped with a headphone attempts to obtain responses, is practiced in industrialized nations. During and after a debate, discussion, or presentation on topics of public interest, soliciting responses through the Internet, e-mail or directly on the television is rapidly becoming popular.

Sample sizes

For a survey on the families in a region, the number of households to be selected in the sample has to be determined before the start of the survey. For a multistage survey, sample sizes for all the stages should

be specified in advance. **Precision** of an estimator, which is defined in Chapter 2, will be high if the variation among all the possible estimates that can be obtained from a probability sampling procedure is small. To determine the size for a planned sample, importance should be given to the required precision, error of estimation, costs of sampling, and the financial resources as well as the time available for the survey.

1.6 Probability sampling

All the sampling procedures described in the following chapters are of this type. Samples selected through probability sampling are also frequently known as **random samples**. Randomness, however, does not imply that the sample units are selected haphazardly. On the contrary, in this method of sampling, the probabilities for selecting the different samples from a population are specified, and the probabilities for the units of the population to appear in the samples are known.

Sampling error is a measure of the departure of all the possible estimates of a probability sampling procedure from the population quantity being estimated. A very important feature of **probability sampling** is that in addition to providing an estimate of the unknown population quantity, it enables the assessment of the sampling error of the estimate, the **standard error**. Further, the estimate and its standard error can be used to obtain **confidence limits** for the unknown population quantity. For a specified probability, these limits provide an interval for the population quantity. Improved procedures of probability sampling to suit practical situations and to estimate population quantities with the desirable properties of high **precision and accuracy** will be presented throughout the following chapters.

1.7 Nonprobability sampling

Purposive sampling

To estimate the total sales or average prices of computers in a metropolitan area, a handful of selected computer stores may thought to be adequate. To estimate the agricultural production in a village, a sample consisting of a certain number of small, medium, and large farms may thought to be **representative**. Similarly, some specified households may be selected for a survey on the family finances in a region. In these

illustrations, prior information on the population units is frequently utilized for selecting the sample.

Quota sampling

In this method, the survey is continued until a predetermined number of the people, households, hospitals, corporations, and similar population units with specified characteristics are contacted and interviewed. In a political survey, for example, the interviewers may be asked to obtain 200 responses from each of the male and female groups between the ages of 25 and 65. In several surveys, people are contacted until specified percentages of responses are obtained to the different items in the questionnaires.

Other types of surveys

In what are known as marketplace mall surveys, interviewers attempt to elicit responses from the shoppers regarding their preferences for consumer products and political candidates, and the results of these surveys are projected to the general public. Fashion magazines elicit responses to questions on the tastes, preferences, attitudes, and opinions of the readers on various topics of interest and extrapolate the results to the entire population.

In some situations, without a proper sampling procedure, it may become convenient to obtain the required information from the readily available population units. Estimates and predictions from these types of nonprobability sampling procedures can sometimes be close to the population values, but their success cannot be guaranteed. Further, the errors of these estimates and predictions cannot be estimated from these types of samples, unlike in the case of probability sampling procedures.

1.8 Sampling in practice

Different procedures of probability sampling are employed suitable to the practical situations. Some of the typical surveys and procedures conducted in practice are presented in the following subsections with brief descriptions of the terminology and procedures of probability sampling. Illustrations of these procedures will be presented throughout the coming chapters.

Demographic and economic surveys

The U.S. Bureau of the Census conducts the Decennial Census of Population and Housing and also a number of ongoing demographic, economic, and household surveys. The bureau also provides services to other government agencies and conducts research on various statistical procedures related to censuses and surveys.

For the Current Population Survey (CPS), conducted every month by the above organization, primary sampling units (PSUs) are first constructed from counties or groups of counties. The PSUs are then grouped into strata with similar characteristics and of approximately equal sizes. At the final stage, about four housing units, the ultimate sampling units (USUs) are selected. For some of the household surveys, the CPS follows a **rotation sampling** scheme. In this procedure, the households selected in the initial sample are included in the sample for 4 months, left out for 8 months, and included again for a further 4 months.

Educational surveys

In several countries, surveys are conducted to examine the educational programs in schools and colleges. Griffith and Frase (1995) describe some of the **longitudinal surveys** of the U.S. Department of Education, which are repeated over periods of time. Assessing the progress and achievements in the schools and colleges and examining the "quality, equity and diversity of opportunity" are the major objectives of these studies. A sample of 8000 students was selected in 1989–90 for the longitudinal study on students completing high school, and a sample of 11,000 was selected for the study on students completing college. The probability samples were selected in two stages. Schools were selected first and then students were chosen from the selected schools.

Employment and expenditure surveys

During the four quarters in a year, the U.S. Bureau of Labor Statistics conducts surveys on the employment and unemployment rates of people, labor conditions, wages, industrial production, technological growth, consumer expenditures, and related topics.

Jacobs et al. (1989) describe the Consumer Expenditure Survey conducted by this bureau. Two ongoing surveys are conducted to obtain information on family expenditures and living conditions and for

providing information for the **Consumer Price Index**. The entire nation is divided into 104 geographical areas, 88 urban and 16 rural, and probability samples are selected from each area. In the Interview Survey, a sample of about 9000 families are contacted and approximately 5000 completed interviews are obtained. The responses are elicited on expensive items like property, rent, automobiles, appliances, utility bills, and insurance premiums. The Diary Surveys are completed at home by a sample of 8000 families on characteristics difficult to recall, such as the expenditures for food, medicine, household supplies, and services.

Health and nutrition surveys

In the U.S., the National Center for Health Statistics conducts a number of surveys on the health and nutritional conditions of inhabitants. Massey et al. (1989) described these surveys for collecting information on the number of visits to hospitals and physicians, short-term and long-term disabilities, chronic conditions, and related characteristics. Approximately, 132,000 persons from a sample of 49,000 households are personally interviewed to obtain the required information.

For these surveys, the entire country is divided into 1900 PSUs. A PSU consists of a county, a small number of contiguous counties, or a metropolitan statistical area (MSA). These PSUs are grouped, **stratified**, according to demographic and socioeconomic variables such as age, sex, race, education, income, and the like. Each PSU is divided into area segments, and each segment is divided into clusters of about eight households. The country is divided into approximately 800 clusters. The actual sampling is conducted in two stages. At the first stage, a sample of PSUs is selected from each stratum with probability proportional to its size. The largest PSUs are included with certainty. At the second stage, a sample of clusters is selected and the households in the selected clusters are interviewed.

Brackstone (1998) describes the **longitudinal survey** to determine the factors affecting Canadian children's physical, social, and cognitive characteristics. A sample of 23,000 children from birth to 11 years of age is selected from 13,500 households. Measurements on the above characteristics are obtained every 2 years on each child until age 25. "In each wave, a new sample of 5000 children" from birth to 23 months of age are added to the sample.

Sampling procedures for the World Fertility Survey was described by Verma et al. (1980). Demographic and health surveys conducted in the developing countries through the cooperation of the UN and WHO

are described by Verma and Thanh (1995). For these surveys, the sampling area in each country is stratified according to the urban or rural type and by the location. At the first stage of sampling, a sample of area units is selected with probability proportional to their sizes. At the second stage, a sample of households is chosen from the selected area units. From women between 15 and 49 years of age in the chosen households, information is obtained on their ages at marriage, reproductive history, number of children, immunization of the children, and other demographic and health-related characteristics.

The Health and Nutrition Survey in Croatia on children 5 years old and younger, conducted in 1996 by the UN International Children's Educational Foundation (UNICEF) was described by Dumicic/ and Dumicic/ (1999). The region unaffected by the war, "front line areas," and "liberated areas" formed the three strata. Each stratum was divided into area segments. In each segment, clusters of 40 households with children under 5 years old were considered. The segments consisted of 120 to 250 households. At the first stage, samples of segments were selected from each stratum with probabilities proportional to the sizes of the segments. At the next stage, samples of clusters were selected from the chosen segments. This procedure resulted in the selection of a sample of 14,800 households consisting a total of approximately 2000 children under age 5.

Edler et al. (1999) describe the Heidelberg Children Health Survey conducted in 1996 on 3828 fifth-grade students. It was repeated in 1998 on 4036 seventh-grade students in 172 classes in 68 schools of the lower, middle, and higher levels of education. The questionnaire with 38 items related to the health of young people was presented to the selected students. The above authors presented data on the smoking habits of the students and their parents, and the frequency of their resulting headaches. It was found that children in families where one or both parents smoke suffered from headaches more frequently than nonsmoking families. There were also significant differences in the smoking habits of children in the above three types of schools.

Agricultural surveys

These surveys are conducted to estimate the acreage and production of crops, livestock numbers, expenditures for agricultural production and labor, utilization of fertilizers and pesticides, and related topics.

Kott (1990) summarizes six surveys conducted by the U.S. Department of Agriculture during 1989–90. In each of the 50 states, farms

of similar sizes are grouped into strata. Samples of farms are selected from the frames, the lists of the farms in the strata. Samples are selected from lists available separately for crops, farm animals, and farm labor. Area frames are also used for the above surveys. In this procedure, each state is divided into segments and a sample of segments is selected. The required information is obtained from the tracts of farm land utilized for agricultural production.

For the agricultural surveys in India described by Narain and Srivastava (1995), each state is stratified into administrative regions. At the first stage, a sample of villages is selected from each region. At the second stage, a sample of agricultural holdings is chosen from each of the selected villages and the required information is obtained.

Implementation of the surveys in Lesotho and Sudan were described by O'Muircheartaigh (1977). Lesotho is a small mountainous country with many rivers and streams, and the ecological zones and population densities were used for stratification. For the agricultural surveys, administrative districts, ecological zones, and the number of cultivated fields were used for stratification. Sudan is a very large and thinly populated country with a vast amount of desert land. The administrative structure within the provinces, rural councils, and such are used for implementing the surveys.

For the agricultural surveys in several countries, each cultivated region is stratified into blocks. At the first stage, a sample of villages is selected from each block. At the second stage, a sample of land holdings is chosen from each of the selected villages and the required information is collected. For the multistage surveys conducted by the UN in Ghana, Kenya, Nigeria, and other African countries to obtain information on agricultural, demographic, and economic characteristics, EAs are selected at the first stage. Landholdings or dwellings are selected at the second stage.

Carfagna (1997) reviews surveys in the European Union conducted through high-resolution satellite images. The entire area is divided into square **sites**, and each site is divided into square segments. A sample of sites is selected first, and then a sample of **segments** is chosen from each of the selected sites. At the next stage, a sample of **points** is chosen from each selected segment and the different crops are estimated.

Marketing surveys

Almost every major company that manufactures or sells consumer products or provides services to the public conducts surveys on the satisfaction and opinions of the consumers. Dutka and Frankel (1995)

describe market surveys in 19 countries on persons 12 to 49 years of age on the brand-awareness of soft drinks, their attitudes toward the brands, and consumption habits. Area sampling is used for these surveys. Each region is **stratified** into urban, rural, metropolitan, and similar divisions. Each stratum consists of PSUs—counties, cities, and similar units. At the first stage, a sample of PSUs is selected. At the second stage, a sample of blocks is chosen from each of the selected PSUs. In the households of the chosen blocks, interviews are attempted on ten people, five males and five females.

Surveys on retail stores in Latin America conducted by the Nielson Marketing Research were described by Santos (1995). The purpose of these surveys was to measure the activity in the stores such as counting of the stocks and purchases, recording the prices, and promotional activities. For these surveys, stratification was based on the geographical region, size of the city, and type of the store, for example, pharmacies and kiosks, and also by the size of the stores determined by the number of employees, selling area, and related information.

Election surveys

In every democratic country, surveys and polls are conducted on the opinions of the people regarding the candidates contesting for public offices such as the president, governors, senators, representatives, and mayors, and the issues of importance at the time of the elections. They are also conducted to forecast the election outcomes ahead of the complete counts of the ballots. In the U.S., television and newspaper networks and several national and local polling organizations conduct such surveys. Brown et al. (1999) describe the procedures followed for the forecasts for the four political parties at the 1997 British elections; the results were used by the British Broadcasting Corporation (BBC).

Public polls and surveys

The news media and several agencies conduct polls and surveys on the current topics of interest to the public. Presenting the results of a national poll, *The New York Times* (Tuesday, February 24, 1998) described how it was conducted. Interviews were obtained via telephone from 1153 adults in the U.S. Telephone exchanges were randomly selected from more than 42,000 residential exchanges across the country. To access both listed and unlisted numbers within each

exchange, random digits were added to form complete telephone numbers. Within each selected household, one respondent was selected randomly to answer the questionnaire. The results were weighted to take account of household sizes and the number of telephone lines into the households and also to adjust for the variability in the sample geographical region, sex, race, age, and education.

In another survey conducted by *The New York Times*/CBS News Poll (*New York Times*, Wednesday, October 20, 1999), 1038 teenagers, 13 to 17 years old, were randomly selected from the 42,000 residential exchanges. The topics included their desire to get good grades, usage of computers and e-mail, need for having friends, problems with drugs and peer pressure, and concerns with violence and crime as well as with their safety.

Campus surveys

College campuses frequently conduct surveys on students regarding their educational programs, living quarters, dining facilities, sports activities, and similar topics of interest. Four of the surveys conducted during the sampling course at the University of Rochester on approximately 4000 undergraduate students are briefly described below.

As a project in the sampling course one year, a survey on the athletic facilities on the campus was conducted. All the students in the university were stratified according to the class—freshman, sophomore, junior, and senior, and the male–female classification. An overall sample of about 200 students, 5% of the total, was distributed to the eight groups proportional to their sizes. Samples were selected randomly from each of the eight groups, approximately 25 from each group, and information on the participation of the students in different types of sports and their visits to the gymnasium was obtained through telephone interviews. More than one attempt was needed to contact some of the sampled students on the telephone, and some of the contacted students either did not respond to the questions or provided only partial responses. Estimates for the numbers of students participating in different types of sports and other characteristics were made only after making suitable adjustments for the nonresponse. Another survey on the students included general topics of interest to the students related to studies, residence halls, meals, alcohol consumption, belief in God, and religious faith. In a food survey conducted on the campus, free samples of two or three items were available during the lunch hour, and the participants rated their preferences on a 1 to 5 scale.

In another food survey, students expressed their opinions on the quality of the food and the services at the three dining facilities. Inferences from these types of surveys are valid if the responses can be considered to be representative of all the students on the campus.

Exercises

1.1. In an assembly or gathering of the public, sometimes the loudness of the applause or the show of hands is used for assessing the approval of a proposition of general interest. When can such a procedure provide a valid estimate for the entire population?

1.2. Consumers shopping at supermarkets are frequently offered samples of food and beverages and their preferences are analyzed. Can the estimates from such analyses be generalized to the entire community?

1.3. People's choices for household purchases are sometimes noted by contacting the persons passing a central location in a shopping mall, for example, one every 5 minutes. Present two situations suitable for generalizing the results of the above type of information to the entire population.

1.4. The opinions of students on the dining facilities at a university were collected during 1 week at the four dining centers on the campus. A total of 300 students expressed their opinions; 90, 80, 65, and 65 of the freshman to senior classes, respectively. The 4000 students consisted of approximately equal numbers in each class. (a) Are the estimates from this type of survey valid for the 4000 students? (b) Can the responses be generalized separately to each of the four classes? (c) Instead of the above procedure, how would one plan to obtain the opinions from a sample of 300 students?

1.5. To estimate the average expenses of approximately 4000 college students for books, entertainment, and travel, the mean of a 5% random sample from all the students can be considered. Alternatively, the sample can be distributed in the four classes, freshmen to seniors, proportional to their sizes and their means weighted accordingly. Which of these two procedures can provide better estimates?

1.6. The average weight of 30 children can be estimated by the mean weight of a random sample of six of the children. Alternatively, as they stand according to their heights, the mean weight of (a) the shortest and tallest child or (b) the two middle ones may be used for estimating the average weight. Which of these procedures can provide a valid estimate for the average weight of the 30 children?

1.7. The households in a region are numbered from 1 to 300. One can consider a 5% random sample from the households to estimate, for example, characteristics related to their education, health, and income. Alternatively, 1 household from the first 20 can be randomly selected and every 20th household from the selected household can be included in the sample. For which situations would each of these two procedures be recommended?

1.8. As another alternative to the two procedures in Exercise 1.7, the households can be randomly divided into 20 groups with 15 in each, and one of the groups can be chosen randomly. When is this procedure preferred?

1.9. Purchases of several products such as refrigerators, air-conditioners, radios, televisions, bicycles, and sporting goods are accompanied by registration and warranty forms. They contain questions related to family size, ages of the household members, educational and income levels, and so forth. The completed and returned forms are sometimes used to estimate the demographic and socioeconomic characteristics of the population and their preferences for the above types of products. Describe two situations for which these practices provide valid estimates.

1.10. In a *straw poll* of the 400 persons at an election meeting, 200 were in favor of the first candidate, 160 the second candidate, but the remaining 40 did not express any opinion. If the 400 can be considered to be representative of all the registered voters, for which situations can the following procedures be recommended? (a) Ignore the 40 without any opinion and estimate the preferences from the 360 responses, (b) add 20 to each of the candidates and consider 220 and 180 to be in favor of the two candidates, and (c) add (200/400)40 and (160/400)40 to the first and second candidates.

1.11. To estimate the percentages of the different brands of computers purchased in a metropolitan area and the average price paid for each brand, two lists were available: (a) the list of all the purchases from the computer stores and (b) the lists of all the purchases from the computer stores, universities, general stores, and through the mail. The second list has twice as many purchases as the first. Information on the purchased brands and the prices paid were obtained from samples of sizes 100 and 200, respectively, from the two lists, and 30 of the purchases appeared in both the samples. From the information observed in these samples, how would one estimate the percentage purchases for the brands and the average of the prices paid.

1.12. For the information collected through the two samples in Exercise 1.11, how would the difference in the average of the prices for the computer stores and other places be estimated?