quality that they can serve as a basis for informed decisions. The methods available to assure good quality should be known and accessible to all serious survey organizations. Today, this is unfortunately not always the case, which is our primary motive and purpose for writing this book.

1.3 BRIEF HISTORY OF SURVEY METHODOLOGY

Surveys have roots that can be traced to biblical times. Madansky (1986) provides an account of censuses described in the Old Testament, which the author refers to as "biblical censuses." It was very important for a country to know approximately how many people it had for both war efforts and taxation purposes. Censuses were therefore carried out in ancient Egypt, Rome, Japan, Greece, and Persia. It was considered a great indication of status for a country to have a large population. For example, as late as around 1700, a Swedish census of population revealed that the Swedish population was much smaller than anticipated. This census result created such concern and embarrassment that the counts were declared confidential by the Swedish government. The government's main concern was a fear that disclosure of small population size might trigger attacks from other countries.

Although survey sampling had been used intuitively for centuries (Stephan, 1948), no specific theory of sampling started to develop until about 1900. For instance, estimating the size of a population when a total count in terms of a census was deemed impossible had occupied the minds of many scientists in Europe long before 1900. The method that was used in some European countries, called *political arithmetic*, was used successfully by Graunt and Eden in England between 1650 and 1800. The political arithmetic is based on ideas that resemble those of ratio estimation (see Chapter 9). By means of birthrates, family sizes, average number of people per house, and personal observations of the scientists in selected districts, it was possible to estimate population size. Some of these estimates were later confirmed by censuses as being highly accurate. Similar attempts were made in France and Belgium. See Fienberg and Tanur (2001) and Bellhouse (1998) for more detailed discussions of these early developments.

The scientific basis for survey methodology has its roots in mathematics, probability theory, and mathematical statistics. Problems involving calculation of number of permutations and number of combinations were solved as early as the tenth century. This work was a prerequisite for probability theory, and in 1540, Cardano defined *probability* in the classical way as "the number of successful outcomes divided by the number of possible outcomes," a definition that is still taught in many elementary statistics courses. In the seventeenth century, Galilei, Fermat, Pascal, Huygens, and Bernoulli developed probability theory. During the next 150 years, scientists such as de Moivre, Laplace, Gauss, and Poisson propelled mathematics, probability, and statistics forward. Limit theorems and distributional functions are among the great contributions during this era, and all those scientists have given their names to some of today's statistical concepts.

The prevailing view in the late nineteenth century and a few decades beyond was that a sample survey was seen as a substitute for a total enumeration or a census. In 1895, a Norwegian by the name of Kiear submitted a proposal to the ISI in which he advocated further investigation into what he called *representative investigations*. The reason that this development was at all interesting was the same faced by Graunt and others. Total enumeration was often impossible because of the elaborated nature of such endeavors in terms of costs but also that a need for detail could not be fulfilled. Kiear was joined by Bowley in his efforts to try to convince the ISI about the usefulness of the representative method. Kiear argued for sampling at three ISI meetings, in 1897, 1901, and 1903. A decade later, Bowley (1913) tried to connect statistical theory and survey design. In a number of papers he discussed random sampling and the need for frames and definitions of primary sampling units. He outlined a theory for purposive selection and provided guidelines for survey design. It should be noted that neither Kiear nor Bowley advocated randomization in all stages. They first advocated a mixture of random and purposive selection.

For instance, one recommendation was that units and small clusters should be chosen randomly or haphazardly, whereas large clusters should be chosen purposively. Independent of these efforts, a very similar development was taking place in Russia led by Tschuprow, who developed formulas for estimates under stratified random sampling. In the mid-1920s the ISI finally agreed to promote an extended investigation and use of these methods. Details on how to achieve representativeness and how to measure the uncertainty associated with using samples instead of total enumerations were not at all clear, though. It would take decades until sampling was fully accepted as a scientific method, at least in some countries.

Some of the results obtained by Tschuprow were developed by Neyman. It is not clear whether Neyman had access to Tschuprow's results when he outlined a theory for sampling from finite populations. The results are to some extent overlapping, but Neyman never referred to the Russian when presenting his early works in the 1920s.

In subsequent years, development of a sample survey theory picked up considerable speed (see Chapter 9). Neyman (1934) delivered a landmark paper "On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection." In his paper Neyman stressed the importance of random sampling. He also dealt with optimum stratification, cluster sampling, the approximate normality of linear estimators for large samples, and a model for purposive selection. His writings constituted a major breakthrough, but it took awhile for his ideas to gain prominence. Neyman's work had its origin in agricultural statistics, and this was also true for the work on experimental design that was conducted by Fisher at Rothamsted. Fisher's work, and his ideas on random experiments were of great importance for survey sampling. Unfortunately, as a result of a major feud between Neyman and Fisher—two of the greatest contributors to statistical theory of all time—development of survey sampling as a scientific discipline was perhaps considerably impaired.

In the 1930s and 1940s most of the basic survey sampling methods used today were developed. Fisher's randomization principle was used and verified in agricultural sampling and subsampling studies. Neyman introduced the theory of confidence intervals, cluster sampling, ratio estimation, and two-phase sampling (see Chapter 9).

Neyman was able to show that the sampling error could be measured by calculating the variance of the estimator. Other error sources were not acknowledged particularly. The first scientist to formally introduce other error estimates was the Indian statistician Mahalanobis. He developed methods for the estimation of errors introduced by field-workers collecting agricultural data. He was able to estimate these errors by a method called *interpenetration*, which is used to this day to estimate errors generated by interviewers, coders, and supervisors who are supposed to have a more-or-less uniform effect on the cases they are involved with, an effect that typically is very individual.

The concepts of sampling theory were developed and refined further by these classical statisticians as well as those to follow, such as Cochran, Yates, Hansen, and others. It was widely known by the 1940s, that sampling error was not synonymous with total survey error. For example, we have already mentioned Mahalanobis's discovery about errors introduced by field-workers. In the 1940s, Hansen and his colleagues at the U.S. Bureau of the Census presented a model for total survey error. In the model, which is usually called the U.S. Census Bureau survey model, the total error of an estimate is measured as the mean squared error of that estimate. Their model provides a means for estimating variance and bias components of the mean squared error using various experimental designs and study schemes. This model showed explicitly that sampling variance is just one type of error and that survey error estimates based on the sampling error alone will lead to underestimates of the total error. The model is described in a paper by Hansen et al. (1964) and the study schemes in Bailar and Dalenius (1969).

Although mathematical statisticians are trained to measure and adjust for error in the data, generally speaking, they are not trained for controlling, reducing, and preventing nonsampling errors in survey work. A reduction in nonsampling errors requires thoughtful planning and careful survey design, incorporating the knowledge and theories of a number of disciplines, including statistics, sociology, psychology, and linguistics. Many error sources concern cognitive and communicative phenomena, and therefore it is not surprising that much research on explaining and preventing nonsampling errors takes place in disciplines other than statistics. [See O'Muircheartaigh (1997) for an overview of developments across these disciplines.]

In the early developments of sampling theory, bias was seldom a concern other than as a technical issue related to characteristics of the estimator itself, like the "technical bias" associated with a ratio estimator (Chapter 9). Early statisticians were not particularly interested in models of response effects, the interaction between the interviewer and the respondent, the complexity of the task to respond or measure, and the realism in variables (i.e., the extent to which measured variables relate to constructs they are meant to describe). Other disciplines assumed that responsibility. There are, for instance, some very early writings on the effects of question wording, such as Muscio (1917). Formal attitude scales were developed by Likert and others during the period 1920–1950. In the 1940s, extensive academic research was conducted on survey instruments when numerous experiments were carried out to identify the strengths and weaknesses of various questionnaire designs.

O'Muircheartaigh also gives an example of an interesting debate in the survey methods literature concerning the roles of interviewers and respondents. The early view of the interviewer held that information was either available or it was not. When it was, it could easily be collected from respondents. Thus, years ago, the primary technique for interviewing respondents was conversational in nature. In one form of conversational interviewing, the interviewer conversed informally with the respondent without necessarily taking notes at the time of the conversation and summarized the information from the interview later. Another form in use was more formal, with the interviewer equipped with a set of prespecified questions that were asked in order as the interviewer took notes.

Interviewer influences on the responses were usually not a concern. Interviewing was primarily a method used in social surveys that, in those days, were generally not held in high regard, due to the lack of control and standardization. Standardization eventually came into greater acceptance. In 1942, Williams provided a set of basic instructions for interviewers at the National Opinion Research Center (NORC) in the United States. In 1946 a discussant at a conference in the United States identified the ideal interviewer as "a married woman, 37 years old, neither adverse to nor steamed up about politics, and able to understand and follow instructions." To some extent, this image of the interviewer, at least as a woman, has prevailed in some interviewing organizations to this day.

There was very little said about respondents' role in the early writings. Issues that deal with interviewer–respondent interaction were not studied until 1968, when schemes for coding these interactions were presented by Cannell and Kahn. In fact, the respondent was often viewed as an obstacle in the data collection process, and this attitude can also be seen today in some survey programs, especially in some of those that are backed up by laws stipulating mandatory participation. A few historical papers, in addition to those already mentioned, include Fienberg and Tanur (1996), Converse (1986), Kish (1995), Hansen et al. (1985), and Zarkovich (1966).

Many new developments have influenced today's survey work. For instance, we now have a sampling theory using model-assisted methods. An example of a modern textbook reflecting these new sampling methods is that of Särndal et al. (1991). Also, response errors can now be incorporated directly into statistical models, and issues of cognition during the interview continue to be studied. There is a continued interest in trying to understand the response process, and new knowledge has increased our ability to improve data collection modes. The development of new technology include computer-assisted data collection, scanning of forms, and using software that makes it possible to convert complex verbal descriptions automatically into numerical codes.

However, to this day, many of the basic problems associated with survey work remain despite vigorous research efforts. These basic problems include the presence of survey errors, the lack of adequate measurement tools and resources to handle the errors, and the lack of understanding by some survey producers, survey users, and survey sponsors as to how errors affect survey estimates and survey analyses. There is need for improved quality in survey work.

LANDMARK EVENTS IN THE HISTORY OF SURVEYS

- The first guidelines for survey design were developed early in the twentieth century.
- Neyman's landmark paper on the representative method was published in 1934.
- In the 1940s, Mahalanobis developed the method of interpenetration of interviewer assignments to estimate errors made by survey field-workers.
- In the early 1960s, Hansen and others developed the first survey model.

1.4 THE QUALITY REVOLUTION

During the last couple of decades, society has witnessed what has been called by its advocates, a *quality revolution* in society. Deming, Juran, Taguchi, Crosby, Ishikawa, Joiner, and others have stressed the need for better quality and how to improve it. For instance, Deming (1986) presented his 14 points and the seven deadly diseases, Juran and Gryna (1980) had their spiral of progress in quality, Taguchi (1986) developed a type of designed experiment where variation is emphasized, Crosby advocated avoiding problems rather than solving them, Ishikawa (1982) listed the seven quality control tools (data collection, histogram, Pareto diagram, fishbone diagram, stratification, plotting, and control charts), and Joiner (Scholtes et al., 1994) emphasized the triangle