Brief History of Statistics

The original meaning of the word "statistics" is science of states, and in its early existence it was also called "political arithmetic". Although the use of the term started as late as the 18th century, the practice of collecting and analyzing data dates back to the early biblical times. At present, the applications of statistics have expanded from the political science to almost all fields of knowledge.

PROBABILITY. Probability theory was inspired mainly by games of chance. The first mathematical analysis of games of chance were undertaken by Italian mathematicians in the 16th century. The main results were initially obtained by Gerolamo Cardano (1501-1576) about 1565. There were many wrong propositions in this book. Although Cardano later realized the errors and eventually corrected them, he didn't remove the false statements in the book. This book was never published because Cardano kept it secret. Accused of being a heretic in 1570, he was arrested, dismissed and denied the rights to lecture publicly and to have his books printed.

The theory of probability had already been discovered by Pierre de Fermat (1601-1665), Blaise Pascal (1623-1662), and Christiaan Huyghens (1629-1695) by the time Cardano's book was rediscovered almost a hundred years later. C. de Mere, a notorious gambler who was a close friend of Pascal, initiated the study of probability theory by mathematicians by posing a gambling problem that offered the mathematicians a real challenge. This led to a lengthy correspondence between Fermat and Pascal, and brought about not only a solution to the problem at hand, but to other more general problems as well. In 1655, the young Dutchman, Huyghens, learned of the results of Fermat and Pascal – not the proofs and arguments, and he wrote a short book on probability theory called "How to Reason in Dice Games". It is fair to say that this book represents the real beginning of probability theory as a mathematical subject.

After a period of stagnation of nearly 50 years, there followed a decade with astounding activity and progress from 1708 to 1718 in which the elementary and fragmentary results of Pascal, Fermat, and Huyghens were developed into a coherent theory of probability. It was at this time that mathematicians discussed elementary rules of probability calculus, conditional probability, expectations, combinatorics, algorithms and recursion formulae, the method of inclusion and exclusion, and examples of using infinite series and limiting process. They derived important special distributions such as the binomial and hypergeometric distributions.

The period that follows was a period of consolidation and steady progress. Many problems were taken up by the following generation of mathematicians and given solutions that have survived until today.

INFERENTIAL STATISTICS. The successful development of probability theory did not immediately lead to a theory of inferential statistics. In 1346, the world faced the most infectious and lethal disease of all times – the black plague. After that time, the plague occurred regularly until 1712. Statistics (or political arithmetic as it was named in its first years of existence) was then the art of deducing estimates and properties of quantities which can not be observed directly. The pioneer seems to be an English tradesman and haberdasher of small wares, John Graunt (1620-1674). In 1661, he published a book called "Natural and Political Observations upon Bills of Mortality". He applied the bills to estimate birth mortality, the number of inhabitants, the number of years to recover the former population level after a plague epidemic, etc. His methods were bold but dubious, but surprisingly were in accordance with later and more reliable observations. Soon the mathematicians involved in the study of probability theory took up the challenge to invent rigorous methods to estimate unknown quantities, particularly to compute reliable life-tables, which became an important tool for the emerging life insurance companies.

The first really significant developments in the theory of statistics as we know it today did not occur until the late nineteenth and early twentieth centuries. The study of methods of making inferences from data began with the work of Francis Galton (1822-1911) and Karl Pearson (1857-1936) in the late nineteenth century. This early period was marked by a change in attitude toward statistics, a recognition of its importance by the scientific world. In addition to this, many advances were made in statistical technique. Among the technical tools invented and studied by Galton, Pearson, and their followers were the standard deviation, correlation coefficient, and the chi-square test.

The second period of development of statistical theory began about 1915 and consists primarily of the work of R. A. Fisher (1890-1962) and his followers. Among their contributions were the development of methods appropriate for small samples, the discovery of the exact distributions of many sample statistics, the formulation of logical principles for testing hypotheses, the invention of the technique known as analysis of variance, and the introduction of criteria for choice among various possible estimators for a population parameter.

The third period began about 1928 with the publication of certain joint papers by Jerzy Neyman and Egon Pearson, the latter a son of Karl Pearson. These papers introduced and emphasized such concepts as 'Type II' error, power of a test, and confidence intervals. It was during this period that industry began to make widespread application of statistical techniques, especially in connection with quality control. There was increasing interest in taking of surveys with consequent attention to the theory and techniques of taking samples.

The fourth period begins in 1939 with the work of Abraham Wald. Perhaps Wald's most significant contribution was his introduction of a new way of looking at statistical problems, what is known as statistical decision theory. From this point of view, statistics is regarded as the art of playing a game, with nature as the opponent. This is a very general theory, and, while it does lead to formidable mathematical complications, it is fair to say that a large share of present-day research statisticians have found it advantageous to adopt this new approach.

The invention of electronic computers in the 1940s and the vast increases in computing power which have taken place since then have had a profound impact on statistics. Computers have made it practical to use techniques which would have been unthinkable without them, and to apply them to previously unthinkably large sets of data. The amount of calculation required by these techniques and the analysis of large sets of data were simply impossible before the advent of electronic computers.

Statistical theory and practice continues to be an important area of research. The study of such topics as "robustness" and "non-parametric statistics" are two important examples from the second half of the twentieth century. The vast amount of polling which accompanies modern elections is just one example of the extent to which statistics has become a central part of modern life.