

<https://doi.org/10.23913/ride.v12i23.1059>

Reseñas

Statistics as a discipline: A brief look to the past, the present and the future

La estadística como una disciplina: una breve mirada al pasado, al presente y al futuro

A Estatística como disciplina: um breve olhar sobre o passado, o presente e o futuro

Mario Miguel Ojeda Ramírez

Universidad Veracruzana, Facultad de Estadística e Informática, México

mojeda@uv.mx

<https://orcid.org/0000-0001-6161-3968>

Roberto Behar Gutiérrez

Universidad del Valle, Facultad de Ingeniería, México

roberto.behar@correounivalle.edu.co

<https://orcid.org/0000-0001-6472-038X>

Pere Grima Cintas

Universidad Politécnica de Cataluña, Departamento de Estadística e Investigación Operativa,

España

pere.grima@upc.edu

<http://orcid.org/0000-0003-1470-1230>

Abstract

Statistics have become especially important in the information and knowledge age. Professionals and scientists, and also citizens, recognize that it helps in the collection, organization and analysis of data, and that its principles support the interpretation and communication of the results obtained. It is accepted that it is a methodology to obtain knowledge, and likewise a technology, that supports diagnoses, interventions and decision-making in contexts of uncertainty. This essay presents a broad and fair conceptualization of this discipline. It also makes a characterization of its mission and briefly reviews its history, trends and current status. At the end, a brief vision of the prospects for the future is presented.

Keywords: Statistical methodology, History of statistics, Teaching statistics, Information and knowledge society, Applied statistics

Resumen

La estadística ha adquirido especial importancia en la era de la información y el conocimiento. Los profesionales y científicos, y también los ciudadanos, reconocen que ayuda a la recopilación, organización y análisis de datos, y que sus principios apoyan la interpretación y comunicación de los resultados obtenidos. Se acepta que es una metodología para obtener conocimiento, y así mismo una tecnología, que respalda diagnósticos, intervenciones y la toma de decisiones en contextos de incertidumbre. Este ensayo presenta una conceptualización amplia y justa de esta disciplina. También hace una caracterización de su misión y revisa brevemente su historia, tendencias y estado actual. Al final se presenta una breve visión de las perspectivas para el futuro.

Palabras clave: Metodología estadística, Historia de la estadística, Enseñanza de la estadística, Sociedad de la información y el conocimiento, Estadística aplicada.

Resumo

A estatística adquiriu especial importância na era da informação e do conhecimento. Profissionais e cientistas, e também cidadãos, reconhecem que ajuda a recolher, organizar e analisar dados e que os seus princípios apoiam a interpretação e comunicação dos resultados obtidos. Aceita-se que é uma metodologia de obtenção de conhecimento, e também uma tecnologia, que suporta diagnósticos, intervenções e tomadas de decisão em contextos de incerteza. Este ensaio apresenta uma conceituação ampla e justa desta disciplina. Também faz uma caracterização de sua missão e faz uma breve revisão de sua história, tendências e situação atual. Ao final, é apresentado um breve panorama das perspectivas para o futuro.

Palavras-chave: Metodologia estatística, História da estatística, Ensino de estatística, Sociedade da informação e do conhecimento, Estatística aplicada.

Fecha Recepción: Diciembre 2020

Fecha Aceptación: Octubre 2021

Introduction

Statistics is applied in many professions and is an important component of graduate courses in many academic departments. It is difficult to find an academic area of study, or a profession, with no data-processing content. Elementary statistical analysis has become ubiquitous in most technical and social sciences and data compilation is an everyday activity. Design of experiments and observational studies and sampling techniques are prominent in undergraduate courses because they are essential elements of a future professional's technical and scientific background. In order to prepare professionals in different disciplines, the approaches and contents of the statistics courses have to be finely tailored to the context of applications and the use of information and communication technologies (ICT) in which the graduates will be engaged in the future.

Statistics is also an academic discipline and a profession in its own right. It is a science, because a lot of statistical activity is concerned with innovation either directly or through consulting and cooperation. Statistics has a high profile in universities of all ranks, either through specialized departments, or as parts of departments of Mathematics, Economics, Biology, Engineering, schools of Medicine, and the like. It encompasses highly theoretical research on foundational issues and in probability theory and narrowly focused applied research (biostatistics, econometrics, clinical trials, psychometrics, quality control, banking and finance,

climatology, and the like). In general, in all areas where empirical research is carried out.

Applications and theory do not operate in isolation. Many important developments in the theory can be traced to practical problems that have found the established methods deficient, and some landmark applications owe a great deal to recently made progress in statistical theory. The statistics profession has been an integral part of key advances in the design of information systems based on censuses, population registers and surveys. Applications and theory do not operate in isolation. Many important developments in the theory can be traced to practical problems that have found the established methods deficient, and some landmark applications owe a great deal to recently made progress in statistical theory. The statistics profession has been an integral part of key advances in the design of information systems based on censuses, population registers and surveys. The diverse forms of practical and theoretical developments permits to qualify statistics as a technology, but also as a science and an art (Fienberg, 2014).

Statistics contributes to a deeper understanding of the human being as a medical subject, an economic, political and social unit of the society, and as an individual (psychology). It contributes to the quality of life both in its principal aspects and on the periphery, such as in culture and sports. Statistical knowledge is needed to understand the figures, graphs and indicators that are common in the mass media. This understanding must be accompanied by an understanding of the implications at the local, national and global levels. The individuals of this century must have the necessary competencies that allow them to actively participate in social, economic and political decisions, and one of the main is statistical literacy.

Statistics has expanded to become an important force in the modern knowledge-and information-based society. Given its wide-ranging involvement and coexistence with other sciences, it is difficult to formulate its definition that would encompass all its activities and modes of operation. There are many definitions, but most of them are limited, partial or incomplete. It is a science and profession of making decisions in the presence of uncertainty with limited resources (Longford, 2013).

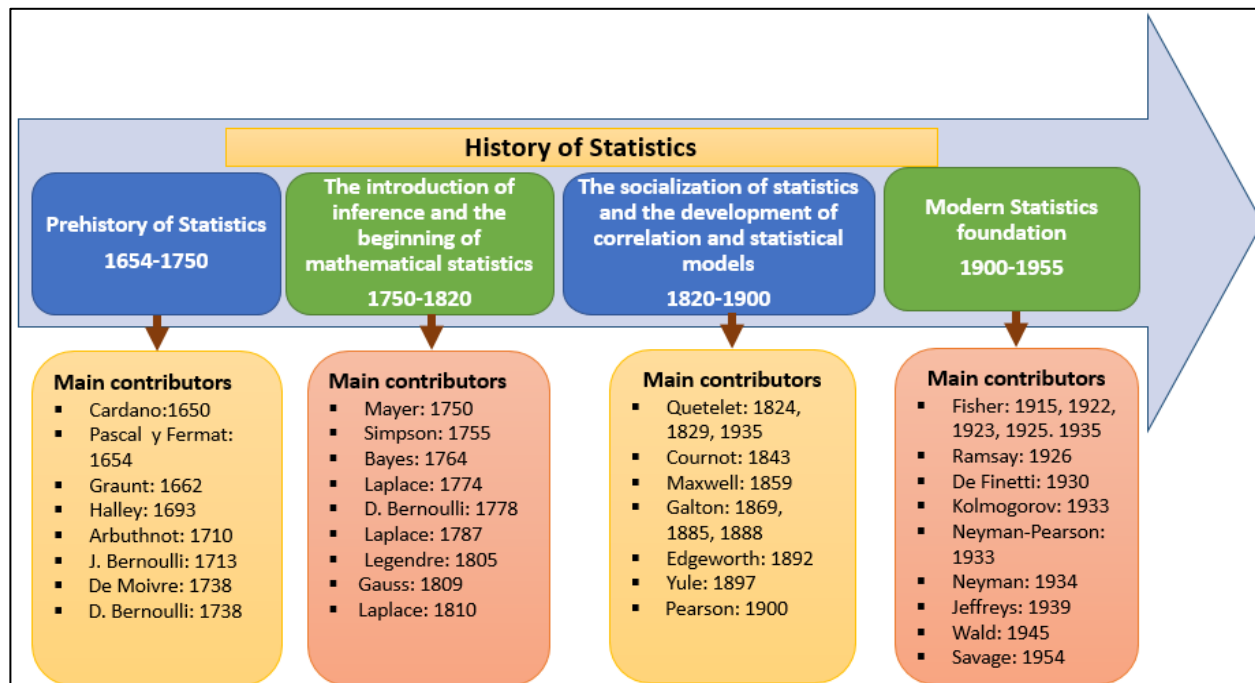
This article presents a comprehensive definition of Statistics. It is developed from a historical perspective and review given in the next section. Some directions in which Statistics may move in the future are explored.

A historical sketch

Statistics has been practiced since the dawn of civilization. The Babylonians, Egyptians, Chinese, Mayans and Incas and all later cultures used to compile and analyze data in the form of counts and quantities, which we now call statistics.

The first formal census was conducted in England in 1066, commissioned by king William I. The emergence of statistics as a scientific discipline set on a systematic foundation is associated with Graunt (1620-1674) who studied mortality in London, and the astronomer Halley (1656-1742), who also contributed to vital statistics. Fienberg (1992), in a famous historical review essay, established the period of the development of probability and the exposition of no probabilistic methods of data analysis as pre-history, and pointed out that the statistical discipline history proper begins around 1750. From this period on development continues in two streams: the socio-demographic and mathematical-encyclopedic. The former culminates in the founding of demography as a discipline, and the latter leads to Statistics in their current form. An important reference about the History of Statistics is Fienberg (1992), who called the period 1750-1820 as *The introduction of inference and the beginning of mathematical statistics*; he reviewed the significant contributions made by Bernoulli (1700-1782), Bayes (1702-1761), Laplace (1749-1827) and Gauss (1777-1855), mainly; continuing the Fienberg's historical review, the period from 1820 to 1900, which was labeled as *The socialization of statistics and the development of correlation and statistical models* and was characterized by the contributions of Quetelet (1796-1874), Galton (1822-1911) and Karl Pearson (1857-1936), between others; they were laid foundations of a theory that supports modern statistics as a scientific discipline and a profession.

Figure 1. Timeline of the development of Statistics as a discipline (adapted from Fienberg, 1992).



Source: Own elaboration using the periods proposed by Fienberg (1992).

These foundations were further developed and set on a firm basis in the 20th century by Fisher (1892-1962), Egon Pearson (1895-1980) Neyman (1894-1981) and Lehmann (1917-2009), between others (see Figure 1). Cox (2016) presents a series of personal reflexions about the work of nine major figures working mostly in the earlier two-thirds of the 20th century, which includes R. A. Fisher, E. S. Pearson, Neyman and: H. Jeffreys (1891–1989), M. S. Bartlett (1901–2001), F. Yates (1902–1984), L. J. Savage (1917–1971), H. E. Daniels (1912–2000) and J. W. Tukey (1915–2000).

R. A. Fisher is considered as the father of modern Statistics. He made important contributions to statistical methodology, motivated by problems in Genetics, Biology and Agriculture. These methods were soon found to be applicable more generally, in industry and social sciences, wherever experimentation and controlled scientific observation is planned. Fisher's reference to statistical method as a 'scientific method' confirms that statistics is a science on par with all others.

The period from 1930s to 1960s saw a rapid expansion of research and application of statistical methodology. Statistics was introduced in research centers concerned with industrial and agricultural production.

A community of professionals in statistics was formed. The subject was introduced in universities, was added to curricula of courses in Agronomy, Biology, Psychology, Economics, Medicine and Engineering. Statistics departments and consulting laboratories were founded.

In the 1950s and 1960s, Statistics became widely recognized, although its application and related research is hampered by the tedious tasks of calculation using primitive equipment. With the advent and proliferation of computers, the techniques of data management and analysis have become an essential part of the social and economic life in the developed countries. From the 1970s on, computers and software implemented on them begin the process of simplifying the application of statistical techniques, moving information from paper-based records to electronic files, making their interrogation fast and flexible and raising the potential of statistical analysis is general.

With this proliferation come a variety of ills, such as disregard for the various assumptions and other misuse of statistical methods. Computer processing time is becoming less and less a concern even for the most extensive operations in data management and iterative inferential algorithms. User-friendly and interactive software with extensive help facilities cuts down the effort needed to master it and promotes their widespread use. The facility with which information is spread in the society and how communication technologies become ubiquitous are changing the paradigm of society of knowledge. These advances in computational capacities have allowed to tackle complex problems, difficult to model with the theory, using a powerful tool that nowadays, it is of extended use: the Monte Carlo simulation. On the other hand, the problem of the restrictions imposed by the assumptions of the statistical methods, have been able to be tackled with nonparametric strategies of intensive computation, like the Bootstrap, of wide acceptance in topicality by the academic community (Efron, 1979).

This brief historical sketch indicates that statistics has developed at a rapid pace throughout the 20th century. The following stepping-stones can be identified in this development:

1. A solid mathematical theory, in particular the axiomatic definition of probability by Kolmogorov (1903-1987), which is the basis of all statistical methodology concerned with management of uncertainty.
2. Personal computers with abundant storage capacity and fast processors and software implemented on them. Dissemination of statistical methods to all areas where information is collected and analyzed. Promotion of methods from the academic and research environment to

all areas of economic activity.

3. A new quantitative paradigm in which statistics plays a central role in establishing scientific validity under uncertainty and supporting the formulation of new knowledge. Presence or impact of statistics in nearly all scientific publications.

The current state

Statistical methodology is the source of tools for five basic tasks:

1, to design studies and conduct them, pointing out research questions for target populations;

2, to compile and document a valid and reliable data bases with minimum cost, effort and delay;

3, to manage database and transform it into format convenient for users;

4, to analyze the data, so that the conclusions are amenable to interpretation that promotes good decisions and assessed the uncertainty and risks arising from sampling variation;

5, to communicate the findings orally or in a written document, or their combination.

This highlights the interdisciplinary nature of much of modern statistics (Fienberg, 2014). As an academic subject, Statistics is concerned with making decisions under uncertainty (Lindley, 1991). This entails design of experiments, surveys and other observational studies, methods for drawing inferences based on them.

In another context, Statistics assists in all economic activities, from production, through retail to public services and administration. Quantitative risk management, an emerging field, presents opportunities for statistical modeling. Many professionals in business and engineer are acquainted with statistical methods and their potential to identify better production processes, more effective provision of services and other goals in a wide range of organizations. It is essential for management and orientation in the vast amount of information generated by rapid technological development. Statistical methods are applied to solve specific problems in more and more diverse areas, and the enclaves where statistics is not used are shrinking all the time.

Statistics has gained a very significant place in society. Below we list some of the factors involved in this process:

1. National and local (regional) governments have statistical systems for planning, decision making and monitoring of economic and social processes. In fact, the UN has a special committee for statistics, which supports member countries in the design and development of such

systems.

2. A wide variety of studies in economics, business and social areas apply statistical methods. They include opinion polling, marketing, banking and finance and many others.

3. The development of medical drugs and devices involves studies conducted according to the statistical principles of experimentation. This process is regulated in most countries. A drug can be released to the market (e.g., for prescriptions) only after evidence about its good properties has been assessed to be sufficient.

4. Quality control in production (manufacture) and the service section is based on statistical methodology.

5. Psychology and educational sciences use Statistics in all the processes they study.

6. Statistical methodology is fundamental to modern life sciences and emerging areas, such as sustainable development.

7. Statistics is introduced in the curricula of primary and secondary education. In some countries, the principles of statistics are disseminated in the general population, promoting universal statistical literacy.

8. Statistics is a recognized profession; 2013 was declared the International Year of Statistics and the UN recognizes a World Statistics Day, October 20, each 5 years from 2010.

9. Several national associations and international institutions, like International Statistical Institute, promote the academic and social activities of statistics.

10. Computer and statistical software implemented on it are ubiquitous and are used with skill by millions of professionals and students. A lot of software, some of it contributed to the established packages, is available freely through the Internet.

An outlook for the future

In the context of the information and knowledge-based society, statistics has a great future (Rao and Székely, 2000; Van Dijk and Hacker, 2003). Its application is getting more and more widespread (Lent, 2002) and its impact is profound. Some obvious directions for its future development are: (1) The science of big data; (2) Increasing complexity of statistical analysis; (3) Further development of hybrid areas of statistics and other sciences: Biometrics, Econometrics, Psychometrics, Cybermetrics and the like; (4) Changes in statistical education; and (5) Promotion of a statistical culture and thinking: statistical literacy.

The availability of big data makes it difficult to draw accurate and useful knowledge to the purposes of understanding complex processes and phenomena. Therefore, the statistical principles accompanied by computer algorithms "of learning and obtaining knowledge" are giving rise to an area, which is expected to be highly dynamic in the coming years: data mining. In fact, to extract information we need Statistics working with several disciplines (Anderson-Cook et al., 2019).

The development of numerical mathematics and statistical computation found a diversity of possibilities to promote computational methods for inference, particularly in the Bayesian approach (Berger, 2002), using simulation and Monte Carlo method for problems of probability which are solved via analytical or would be very difficult or impossible. In this sense, the school of Bayesian statistical inference is occupying a large space for the development of statistical science, and is expected to soon be the dominant approach (<http://www.bayesian.org>).

The concern for promoting of a statistical culture that is, to be part of a general culture of the information and knowledge society, has generated several initiatives in the statistical community, but it is expected that in the coming decades it will be a task more general, which of course will involve professionals of statistics, but especially managers and educational system actors. Also it will be considered mass media professionals. Now, the role of communication of statistics is very important. Statistical literacy; i.e., the ability to understand and critically evaluate statistical displays and results, is a goal in several national programs, in order to build a knowledge society. It is concluded, in this sense, that the major part of the society has to perceive that the level of its statistical culture is not very high and is not quite in line with the standards of the democratic civic society that is why a lot has to be learned and developed.

The 2020 pandemic has shown the need to collect data using homogeneous systems that produce reliable figures and indicators in real time to make decisions every day. It is known that the number of infections and deaths is being underestimated in many countries where health systems are deficient and epidemiological monitoring protocols are lax. Statistics show society the areas in which it must improve, this time with a very regrettable illustration. The focus of the training of statistical thinking (Wild and Pfannkuch, 1999) in different professionals that require this methodology is an issue that occupies increasing attention of researchers in the area of statistics education.

In general, as regards the statistical training of non-statistical professionals, the trend is to remove the calculations, as the center of the teaching and strengthening statistical thinking. On the other hand, to change the approach of teaching tools to answer research questions, considering the whole process, since the approach of the problem and the generation of the data; against the tendency to consider statistics only like data analysis. These changes strengthen the phase of the transit of the real world to the symbolic world as well as in the opposite direction (Batanero, 2019).

Conclusions

Under the actual conception, we expect a growth in the demand of data scientists in the workspaces in new niches, considering the dynamics of development that we have presented. Data scientists work in gathering data, in processing for obtaining useful information, and presenting that information to decision making. Data scientists do not focus on a single source but look at data from multiple sources to enable them to predict trends and forecast and interpret results from the collected data. Several studies indicated that data scientists are experienced professionals with possessing skills on data processing or data visualization, programming language, statistical packages, office applications and databases (Anderson-Cook et al., 2019). Data scientists appear to be closely related to statisticians but also to big data. In the case of developing societies in the immediate future the labor market for professionals in Statistics is much more flattering, as there is a large deficit of these professionals; of course they should be prepared to adapt to the rapid changes that discipline suffer. H. G. Wells did say, “Statistical thinking will on day be as necessary a qualification for efficient citizenship as the ability to read and write”. The statistic is just now part of a new culture that emerges: the called information and knowledge society, and in this sense, statistics should expand all the influences. The future is here.

Future lines of research

The history of statistics has run its course; however, after the essay by Fienberg (1992), few are the integrative and synthesizing works that have been published. This essay is located along this line, from which several challenges arise. The first is to make a detailed analysis of the main events after the 1950s. Here, without a doubt, the role that the computer and statistical packages have played is decisive. This will lead to the identification of aspects such as business

intelligence and data mining, leading to "data science"; It is also important to locate the role of the Quality Revolution and the development of industrial and business statistics. Rao & Székely (2000) and Tanner & Wells (2000), among others, present reviews, from different areas of the discipline, of the perspectives of statistics for the 21st century, which is a basis for undertaking a synthesis work. Many journals include works on the history of statistics, some even at the section level, such as *The American Statistician*, which indicates the relevance of this topic.

Acknowledges

This work was carried out while the authors met in Veracruzana University, in several conferences during six years ago. We acknowledge the support from Nick Longford for his careful reading of several previous versions, and his comments and useful suggestions that have improved the quality and the readability of the paper. We also thank the anonymous reviewers for their comments and suggestions for improving our review and the expository strategy.

References

- Anderson-Cook, C. M., Lu, L., & Parker, P. A. (2019). Effective interdisciplinary collaboration between statisticians and other subject matter experts. *Quality Engineering*, 31(1), 164-176.
- Batanero, C. (2019). Thirty years of research in stochastic education: Reflections and challenges. [Treinta años de investigación en educación estocástica: Reflexiones y desafíos.] In J. M. Contreras, M. M. Gea, M. M. López-Martín & E. Molina-Portillo (eds.). *Actas del Tercer Congreso Internacional Virtual de Educación Estadística*. www.ugr.es/local/fqm126/civeest.html
- Berger, J. O. (2002). Bayesian analysis: a look at today and thoughts of tomorrow. In A. E. Raftery, M. A. Tanner & M. T. Wells (eds.). *Statistics in the 21st Century*. (pp. 275-290). Chapman and Hall.
- Cox, D. R. (2016). Some pioneers of modern statistical theory: a personal reflection. *Biometrika*, 103(4), 747-759.
- Efron B. (1979). Bootstrap methods: Another look at the Jackknife. *The Annals of Statistics*, 7, 15-30.
- Fienberg, S. E. (1992). A brief history of statistics in three and one-half chapters: A review essay. *Statistical Science*, 12(2), 208-225.

- Fienberg, S. E. (2014). What is statistics? *Annual Review of Statistics and its Applications*, 1, 1-9.
- Longford, N. (2013). *Statistical decision theory*. Springer Verlag.
- Rao, C. R. & Székely, G. J. (eds.) (2000). *Statistics for the 21st century*. Marcel Dekker.
- Tanner, M. A., & Wells M. T. (eds.). (2002). *Statistics in the 21st Century*. (pp. 275-290). Chapman and Hall.
- Van Dyk D. A. (2014). The role of statistics in the discovery of a Higgs Boson. *Annual Review of Statistics and its Application*, 1, 41-59.
- Wild, C. J. & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, 67(3), 223-265.

Rol de Contribución	Autores
Conceptualización	Mario Miguel Ojeda Ramírez (Principal), Roberto Behar Gutiérrez (Igual) y Pere Grima Cintas (Igual)
Escritura - Preparación del borrador original	Mario Miguel Ojeda Ramírez (Principal), Roberto Behar Gutiérrez (Apoya) y Pere Grima Cintas (Apoya)
Escritura - Revisión y edición	Mario Miguel Ojeda Ramírez (Principal), Roberto Behar Gutiérrez (Apoya) y Pere Grima Cintas (Apoya)
Adquisición de fondos	Mario Miguel Ojeda Ramírez