Received: 3 May 2022 | Revised: 7 March 2023 | Accepted: 12 March 2023

DOI: 10.1111/sjos.12641

ORIGINAL ARTICLE

Scandinavian Journal of Statistics

A historical overview of textbook presentations of statistical science

Alan Agresti

Department of Statistics, University of Florida, Gainesville, Florida 32611, USA

Correspondence
Alan Agresti, Department of Statistics, University of Florida, Gainesville, FL 32611, USA.
Email: agresti@ufl.edu, aa@stat.ufl.edu

Abstract

We discuss the evolution in the presentation of statistical science in English-language textbooks, focusing on the period 1900–1970 as the field became increasingly influenced by research contributions of R. A. Fisher and Jerzy Neyman. George Udny Yule authored an early popular book that had 14 editions. Methods books authored by Fisher and George Snedecor guided scientists in implementing modern statistical methods. In the World War 2 era, textbooks authored by Maurice Kendall, Samuel Wilks, and Harald Cramér presented a dramatically different "mathematical statistics" portrayal that centered on theoretical foundations. The textbook emergence of the Bayesian approach occurred later, influenced by books by Harold Jeffreys and Leonard J. Savage. The quarter century after World War 2 saw an explosion of books in mathematical statistics and in particular topic areas. In addition to his highly cited research contributions, Sir David Cox was a prolific author of books on a great variety of topics. Most were published after the 1900–1970 period considered in this article, but we also summarize them as part of this special issue to honor his memory. We conclude by discussing the future of textbooks on the foundations of statistical science in the emerging, ever-broader, era of data science.

KEYWORDS

Bayesian statistics, David Roxbee Cox, George Snedecor, George Udny Yule, Harald Cramér, history of statistics, mathematical statistics, Maurice Kendall, R. A. Fisher, Samuel Wilks

1 | INTRODUCTION

This article discusses the historical presentation of statistical science in textbooks. We focus on the first 70 years of the 20th century, the period during which statistical science had the great advances that define its foundations today. We focus on two periods: Before World War 2, the research contributions of R. A. Fisher revolutionized the field from the era of Karl Pearson and started to be noted in books on statistical methods but had limited visibility in the few textbooks that dealt with statistical theory. The second is from World War 2 through the 1960s, wherein many universities started Statistics departments, and textbooks on statistical theory had a more mature focus, including the Fisher and Neyman–Pearson results. Our discussion is limited to English-language textbooks. With our 1900–1970 focus, we also do not discuss books published well before then that contained elements of modern statistical science, such as Pierre–Simon Laplace's 1812 *Théorie Analytique des Probabilités*.

George Udny Yule (1911) published *An Introduction to the Theory of Statistics*, the first in-depth presentation of statistical science as developed in the Galton/Pearson era. The many editions of Yule's textbook were the primary textbook source on statistical science foundations for about 30 years. Section 2 discusses Yule's and other early contributions from Britain, and Section 3 covers corresponding publications during this period in the United States. Section 4 then discusses two books that were highly influential before World War 2 in introducing modern statistical methods, in particular Fisher's innovations: *Statistical Methods for Research Workers* by Fisher (1925) and *Statistical Methods* by Snedecor (1937).

During and immediately following World War 2, a very different type of *mathematical statistics* textbook appeared. Section 5 discusses books authored by Samuel S. Wilks, Maurice G. Kendall, and Harald Cramér that helped to establish statistical science as the mature field of today that has the Fisher and Neyman–Pearson underpinnings. Section 6 summarizes the explosion in such textbooks that occurred in the following quarter century. Section 7 discusses key books in primary subject-matter areas published in that quarter century as many universities established Statistics departments. Section 8 discusses books that advocated the Bayesian approach to probability and statistical inference. Sir David Cox was a prolific author of books on an impressive variety of topics, from 1949 throughout his career. Most of these books were published after the period 1900–1970 of focus in this article, but Section 9 summarizes them as a way of honoring him in this special issue. The final section briefly considers the future of textbooks in statistical science, as the field further evolves with increasing types of data and methods as part of the encompassing field of data science.

A companion article (Agresti, 2021, available at https://www.stat.ufl.edu/∼aa/articles/Agresti_BJPS.pdf) has more detail about statistical theory textbooks, including showing the Table of Contents of the books discussed in this article by Yule, Rietz, Fisher, Snedecor, Wilks, Kendall, Cramér, Mood, Hogg and Craig, Fisz, Rao, Jeffreys, Savage, Schlaifer, and Raiffa and Schlaifer. The article also includes comments from several notable statisticians, including David Cox, about the aspects of statistical science that should now be presented in textbooks on the foundations of the subject.

2 | EARLY 20TH-CENTURY BRITAIN: BOWLEY AND YULE STATISTICS TEXTBOOKS

Bowley (1901) published *Elements of Statistics*, based on lectures that he had presented at the London School of Economics and Political Science since its foundation in 1895. Bowley's book presents innovative ways to summarize the data graphically and numerically. However, its contents illustrate how the essence of the field would soon change. It shows SE formulas for means and differences of means, but otherwise does not consider statistical inference. The final 10 pages introduce the correlation, regression toward the mean, and regression equations and their importance in the study of evolution, but with no discussion of handling multiple variables, such as in multiple regression. The book was popular and had seven editions, the final one appeared in 1937.

In 1911, the year in which Karl Pearson founded the first academic Statistics department at University College, London, George Udny Yule published *An Introduction to the Theory of Statistics* (Yule, 1911). Yule had begun working at University College in 1893 as a demonstrator for Karl Pearson, and he lectured on Statistics there from 1902 until 1912. His research contributions included introducing the partial correlation and measures of association for contingency tables. For such tables, he showed the potential discrepancy between marginal and conditional associations, illustrating this by the phenomenon now called *Simpson's paradox*. In promoting the odds ratio, Yule was critical of Pearson's approach of assuming that continuous distributions such as the bivariate normal underlie two-way contingency tables. For time series analyses, he later devised the correlogram and laid the foundation for autoregressive modeling. Yule's textbook, which introduced methods of the Galton/Pearson era as well as Yule's own ideas in a relatively nontechnical manner, portrays well the state of statistical science in Britain in 1911. Compared with Bowley's book, Yule's has higher mathematical level and a much wider variety of topics. For instance, besides material on descriptive measures and standard distributions, Yule's book contains substantial material on correlation and on conditional associations, such as in discussing multiway contingency tables and deriving partial regression coefficients for multiple regression models and the corresponding partial correlation measure. The exercises deal with theory and methods as well as analyses of real data.

Reviews of Yule's book were generally complimentary, but because Yule had proposed a different approach to modeling association with discrete data, Pearson himself was highly critical. In a 1913 article with D. Heron in *Biometrika*, he stated "If Mr. Yule's views are accepted, irreparable damage will be done to the growth of modern statistical theory. … Unthinking praise has been bestowed on a text-book which at many points can only lead statistical students hopelessly astray." When Yule's health did not permit preparing new editions, a chance meeting with Maurice G. Kendall resulted in Kendall becoming a co-author, from the 11th edition in 1937 until its final 14th edition in 1950. Kendall added chapters on moments and measures of skewness and kurtosis, the chi-squared distribution, small-sample methods, the analysis of variance, and time series. Perhaps surprisingly, none of the editions mention maximum likelihood or R. A. Fisher's other important contributions in the 1920s on statistical theory and the design of experiments, or the Neyman–Pearson theory of hypothesis testing.

3 | THE FIRST MATHEMATICAL STATISTICS TEXTBOOKS IN THE UNITED STATES

In 1918 Carl J. West, a mathematics professor at Ohio State University, published *Introduction to Mathematical Statistics*, perhaps the first occurrence of "mathematical statistics" in a book title (West, 1918). However, the book uses only elementary mathematics and its exercises focus on calculations rather than theory. Topics include descriptive statistics and graphics, correlation, association in contingency tables, and Pearson curves.

In 1927 Henry Lewis Rietz of the University of Iowa provided a stronger theoretical presentation in his textbook, *Mathematical Statistics* (Rietz (1927)). Rietz was a founding member of the Institute of Mathematical Statistics in 1935 and its first president. Topics include the relative frequency definition of probability, mathematical expectation, moments of random variables, the binomial distribution and its normal and Poisson approximations, the Pearson estimate of correlation, linear regression with least squares estimates and the bivariate normal distribution, multiple correlation and partial correlation, and the SE for a variety of measures. It briefly mentions the 1908 Student (pseudonym for Gosset) paper for dealing with small samples, but fails to discuss statistical inference or Fisher's landmark ideas about likelihood and related concepts. In retrospect, the books by West and Rietz were not landmarks, but it would be nearly another 20 years before mathematical statistics books appeared that presented the concepts of statistical inference that are now regarded as fundamental.

4 | FISHER AND SNEDECOR TEXTBOOKS ON STATISTICAL METHODS

Ronald Aylmer Fisher's work at Rothamsted Experimental Station beginning in 1919 and his fundamental research contributions in the early 1920s led naturally to his 1925 book, *Statistical Methods for Research Workers*. This was the first book that discussed maximum likelihood, t tests, the analysis of variance, and randomization and blocking in the design of experiments. Topics include an explanation of what is meant by consistent, efficient, and sufficient statistics and the method of maximum likelihood. The book points out Karl Pearson's error in stating the degrees of freedom for the chi-squared test, shows how to apply the t distribution to significance testing of the mean with small samples, and contains the first textbook example of the analysis of variance for an actual experiment. Tables at the end of the book include t distribution quantiles for degrees of freedom values between 1 and 30.

Harold Hotelling favorably reviewed Fisher's book in *Journal of the American Statistical Association* in 1927, stating "The author's work is of revolutionary importance and should be far better known in this country." He also reviewed the next six editions of Fisher's book for the journal and wrote an article about Fisher's influence to celebrate the 25th anniversary of the publication of the first edition (Hotelling, 1951). The 14th and final edition was published in 1970, completed by E. A. Cornish based on notes that Fisher made before his death in 1962. However, all editions of Fisher's book were difficult reading for most scientists who analyzed data.

Snedecor's (1937) book *Statistical Methods* was more user-friendly than Fisher's and probably had a greater impact on increasing their use. Appointed to the mathematics department at Iowa State College in 1913, Snedecor taught Statistics and formed the Mathematics Statistical Service in 1927 and the Statistical Laboratory in 1933. Snedecor was an early admirer of Fisher's contributions, inviting him to give summer courses at Iowa State in 1931 and 1936 and awarding him his first honorary doctorate in 1936. In Snedecor's book, statistical inference appears right from the start, with the third page distinguishing between sample and population and the following page introducing hypotheses about parameters and then a chi-squared test of goodness of fit. Several chapters deal with regression, correlation, and analysis of variance and covariance. The influence of Fisher's 1935 book on *Design of Experiments* is shown by the great attention that topic also receives. One helpful chapter presents suggestions about methods to use with calculating machines. In a 1974 obituary of Snedecor in *International Statistical Review*, Oscar Kempthorne recalled that "the great majority of workers in noisy sciences loved the Snedecor 'cookbook.'

…they could feel sympathetic to the examples, they could see how statistical methods advanced the quality of the picture, they could see how the same methods could be applied to their own problems, and finally they could *feel* that the process of following the Snedecor cookbook had improved their ability to interact constructively with their data. …Snedecor's presentation is the exemplar to be surpassed."

The depth of explanation in Snedecor's book improved when the 5th edition in 1956 added William Cochran as co-author, who had conducted significant research in experimental designs and surveys. While a postgraduate student at the University of Cambridge, Cochran was hired in 1933 at Rothamsted by Frank Yates, after Fisher left Rothamsted for University College. Cochran himself left Rothamsted in 1939 for Iowa State, where he stayed until taking positions after World War 2 at North Carolina State, Johns Hopkins, and Harvard. New topics added in the 5th edition include multiple comparisons, the Wilcoxon test, and expectations of mean squares in factorial ANOVA. The 8th and final edition of Snedecor and Cochran appeared in 1989.

5 | THE WORLD WAR 2 PERIOD: WILKS, KENDALL, CRAMÉR, AND MODERN MATHEMATICAL STATISTICS

Hotelling (1940) stated "A knowledge of theoretical statistics is not easy to obtain. There is no comprehensive treatise on the subject, starting with first principles, and proceeding by sound deductions and well-chosen definitions to the methods that need to be used in practice. …the seeker after truth regarding statistical theory must make his way through or around an enormous amount of trash and downright error." That situation was soon to change, due to textbooks of a more theoretical nature written by Samuel S. Wilks, Maurice G. Kendall, and Harald Cramér.

Samuel Stanley Wilks received his PhD at the University of Iowa in 1931, supervised by Henry Rietz. Wilks was a founding member of the Institute of Mathematical Statistics in 1935 and served as the editor of *Annals of Mathematical Statistics* from 1938 until 1949 as the journal become one of the foremost Statistics journals. Appointed to a position in mathematics at Princeton University, Wilks taught a Statistics course, the notes for which evolved by 1943 into a lithoprinted publication, *Mathematical Statistics*, later published in 1947 by Princeton University Press (Wilks, 1947).

The outline of Wilks's book is quite different from earlier books on statistical theory, and similar to how many instructors would present the topic today. After presenting probability as a long-run relative frequency and introducing probability distributions, the book discusses moments and moment generating functions, the Central Limit Theorem, sampling distributions that occur in sampling from a normal population, and the independence of the sample mean and variance with such sampling. Wilks then introduces confidence intervals and maximum likelihood estimation, with discussion of Fisher's principles of consistency, efficiency, and sufficiency. He shows how to generate statistical tests using confidence intervals, pointing out that "confidence intervals give us a far more complete judgement about the parameter than significance tests." He also introduces likelihood-ratio tests, his result about its limiting chi-squared distribution, and the Neyman–Pearson theory and its concepts of a critical region and power. Wilks then applies regression to analysis of variance problems, with discussion of topics such as randomized blocks, Latin squares, incomplete layouts, and the analysis of covariance. Topics in later chapters include tests of independence in two-way contingency tables, Hotelling's multivariate version of the t test and its generalization for testing equality of multivariate means, multivariate regression and analysis of variance, principal component analysis, and canonical correlation theory. Although this book has greater mathematical depth than Yule's, it also has a very different style, being quite dry and sparse in motivation and failing to include examples or data-based exercises. It was favorably reviewed by prominent statisticians, including Jerzy Neyman, Abraham Wald, and Maurice Bartlett. Lehmann (2008) noted "To pull these various contributions together and present them in a unified, coherent account, combined with the necessary mathematical and probabilistic preparation, was a major achievement."

Maurice G. Kendall's *The Advanced Theory of Statistics* also appeared in 1943 (Kendall, 1943). The book was originally planned in 1938 as a joint project with M. S. Bartlett, J. O. Irwin, E. S. Pearson, and J. Wishart, but after the outbreak of World War 2, Kendall had to proceed alone. After introducing descriptive statistics, standard probability distributions, and concepts of probability, the book deals with sampling and sampling distributions and with association and correlation. Kendall added a second volume in 1946 which was strongly influenced by research results of Fisher and Neyman and which presented methods of statistical inference and regression analysis (Kendall, 1946). The material on estimation includes properties of maximum likelihood, Fisher's fiducial approach, and the use of information from the log-likelihood function for large-sample confidence intervals. The discussion of statistical tests presents the t statistic, the use of the F distribution in the analysis of variance, and the Neyman–Pearson approach, with discussion of unbiased tests and likelihood-ratio tests. A chapter on multivariate analysis presents Hotelling's generalizations of the t test, Fisher's linear discriminant analysis, and canonical correlation analysis. Kendall also discusses sampling methods, the design of experiments, and time series analysis using moving averages and autoregressive series. Most of the exercises are highly technical, often presenting recent research results.

The breadth and depth of Kendall's treatment of the foundational results of the previous quarter century is quite astonishing when one considers the circumstances in which he wrote the nearly 1000 pages of these two volumes. An email to me in 2021 from David Cox stated, "Kendall was a man of phenomenal energy. Kendall volume 1 was largely written in air-raid shelters. At the same time MGK had a demanding full-time job in shipping control. World War 2 had the effect of a massive increase in interest in academic statistics, in Cambridge, Oxford, Imperial in the UK and Harvard, Berkeley, Columbia, Stanford, etc. In my own case it is highly unlikely that I would have become a statistician without the War, and I am one of many."

The third landmark publication in statistical theory in this period was Cramér's (1946) *Mathematical Methods of Statistics*. Cramér was director of the Institute of Mathematical Statistics at the University of Stockholm and President of that university from 1950 to 1961. Compared to the other books, Cramér's book has higher mathematical level and much more emphasis on probability. The book's first part covers set theory, Lebesgue measure and integration, and matrix algebra. The second part of the book, on "Random Variables and Probability Distributions," based on a separate tract that he had published in 1937, mentions logical approaches of Jeffreys and Keynes, states Kolmogorov's axiomatic approach, covers probability distributions based on the normal distribution as well as others, presents convergence theorems such as versions of the law of large numbers, introduces characteristic functions and proves the Central Limit Theorem, and discusses the delta method and other generalizations. The book's third part on statistical inference introduces sampling and sampling distributions, tests of significance, analysis of variance, and regression and correlation. Like Wilks's book and Kendall's volumes, the book is up-to-date in presenting the Fisher and Neyman–Pearson advances. It has careful explanation of hypothesis testing, the theory of point estimation including maximum likelihood estimation and its asymptotic properties and concepts of efficiency and sufficiency, and confidence regions, focusing on Neyman's approach. Unlike Wilks and Kendall, Cramér discusses the Bayesian approach. Among other innovations, the book introduces what became known as the Cramér–Rao lower bound for the variance of an estimator.

Cramér's book received highly favorable reviews, particularly for its mathematical sophistication. Lehmann (2008) recalled that the book "quickly established itself as the standard introduction to the theory of statistical inference developed by Fisher and Neyman–Pearson." In his comments about such books in Agresti (2021), David Cox stated "Cramér was indeed a towering figure in Sweden; the English translation of his general book was massively influential in the U.S. in particular."

In 1939 Alexander C. Aitken, a New Zealand native who spent his academic career in Edinburgh, Scotland, published *Statistical Mathematics*, a much shorter book but one that also placed strong emphasis on mathematics (Aitken, 1939). This may have been the first book in which a systematic development of the theorems of mathematical statistics used moment generating functions. Another relatively short book that appeared soon after World War 2 with a mathematical formulation including substantial use of moment generating functions was *A First Course in Mathematical Statistics* by Weatherburn (1946), based on his lectures at the University of Western Australia.

6 | THE QUARTER CENTURY AFTER WORLD WAR 2: EXPLOSION IN MATHEMATICAL STATISTICS

Following World War 2, Statistics departments became more common in universities in North America, particularly with the explosion in the growth of state universities in the 1960s with the baby boomer generation (Agresti & Meng, 2013). Likewise, most universities initiated courses in statistical theory. We next discuss some successors to the textbooks of Wilks, Kendall, and Cramér that were often adopted for such courses during the next quarter century. Having somewhat lower technical level, they were specifically designed as textbooks for a first-year graduate-level or an advanced undergraduate-level course in statistical theory.

An excellent example is Alexander Mood's (1950) *Introduction to the Theory of Statistics*. Mood had studied at Princeton under Samuel Wilks. After World War 2, he was recruited by William Cochran to Iowa State, where he initiated formation of the Statistics department. Mood's book discusses probability, discrete and continuous distributions, expected values and moments, sampling, point and interval estimation, the multivariate normal distribution, tests of hypotheses, regression, experimental design and the analysis of variance, sequential tests, and distribution-free methods. He distinguishes applied mathematics from the practice of statistics: "The use of statistical tools is not merely a matter of picking out the wrench that fits the bolt; it is more a matter of selecting the correct one of several wrenches which appear to fit the bolt about equally well but none of which fit it exactly." Any statistical formula "is merely a tool, and moreover a tool derived from some simple mathematical model which cannot possibly represent the actual situation with any great precision. In using the tool one must make a whole series of judgments relative to the nature and magnitude of the various errors engendered by the discrepancies between the model and the actual experiment." When Mood did not have time to revise his book, McGraw Hill Publishing enlisted Franklin Graybill to prepare a 2nd edition in 1963. The 3rd edition added Duane Boes in 1974.

In 1959, Hogg and Craig (1959) published *Introduction to Mathematical Statistics*, which soon became a popular textbook about statistical foundations. Craig received the first Master's degree in mathematics at the University of Florida in 1928 and then joined the University of Iowa at the same time as Samuel Wilks, attracted by the presence of Henry Rietz. Hogg received his PhD at Iowa with Craig as advisor and was founding chair of the Statistics Department in 1965. Their textbook included the by-now standard topics of probability, discrete and continuous distributions, point and interval estimation, and hypothesis testing, but differed from other books in including greater discussion of sufficient statistics and the change-of-variable method of deriving the distribution of a function of several random variables. Early editions, however, have few pages on regression and no mention of Fisher except for attributing the factorization criterion for a sufficient statistic to him and Neyman. The Hogg and Craig textbook is still in print, with its 8th edition published in 2018 with Joseph McKean as co-author.

Other textbooks of similar level as those of Mood and of Hogg and Craig were published during the 1950s but did not receive as much course textbook adoption. For instance, *Statistical Theory with Engineering Applications* by Hald (1952) is a translation of a Danish textbook by Hald, a professor of Statistics at the University of Copenhagen. This 783-page summary of probability and statistical theory includes multiple examples, mainly from engineering. However, it is a mathematical statistics book designed for statisticians perhaps dealing with engineering problems rather than for engineers wanting to learn how to apply statistical methods. The book contains the by-now standard topics but also some less common ones, such as the distribution of the range, statistical control, and sampling inspection. Hald also published a companion book of tables and formulas. Another fine example is the 1958 *Statistics: An Introduction* by Donald A. S. Fraser, who was near the beginning of a distinguished career at the University of Toronto (Fraser, 1958). Fraser's book covers the same primary topics as others of the era but also covers some unusual ones, such as in introducing analysis of variance for sampling from a finite population. In the early 1960s, textbooks were published with similar content but having slightly lower technical level and greater potential for use in undergraduate courses. Examples are *An Introduction to Mathematical Statistics* by H. Daniel Brunk (1960), *Mathematical Statistics* by Freund (1962), and *Statistical Theory* by Lindgren (1962).

At a higher technical level than Mood or Hogg and Craig, during this period Maurice Kendall and Samuel Wilks extended their earlier World War 2 era contributions. Kendall added Alan Stuart as co-author for *The Advanced Theory of Statistics*, expanding it from two volumes to three: *Volume 1: Distribution Theory* in Kendall and Stuart (1958), *Volume 2: Inference and Relationship* in Kendall and Stuart (1961a), and *Volume 3: Design and Analysis, and Time Series* in Kendall and Stuart (1961b). Kendall was knighted in 1974. For the most recent 6th editions, volume 1 in 2010 (with title *Kendall's Advanced Theory of Statistics: Distribution Theory*) has J. Keith Ord as co-author and volume 2 in 2015 (with title *Kendall's Advanced Theory of Statistics: Classical Inference and the Linear Model*) has Ord and Steven Arnold as co-authors. The 4th edition of volume 3 appeared in 1982 with Ord as co-author. Samuel S. Wilks (1962) published a revision of his 1943 *Mathematical Statistics* that more than doubled its length. The book includes three chapters on estimation (linear, nonparametric, parametric), two on hypothesis testing (parametric and nonparametric), as well as chapters on sequential analysis and decision theory, time series, multivariate analysis, and linear statistical estimation. A review by David Cox in 1962 in *Annals of Mathematical Statistics* concludes that although the Kendall and Stuart volumes would probably be preferred by working statisticians looking for an introduction to a particular advanced topic, Wilks's book had a choice of topics more in line with conventional interests and was likely to be preferred as a textbook for advanced students. However, Cox also criticized Wilks in failing to examine the main concepts of the subject and connections between confidence intervals and tests.

Another advanced book published soon after Wilks's was Calyampudi Radhakrishna (C. R.) Rao's (1965) *Linear Statistical Inference and Its Applications*. Rao was strongly influenced by his

PhD advisor, R. A. Fisher, and his treatment of statistical inference focuses on the Fisher and Neyman–Pearson schools. The book begins with a chapter on vector spaces and matrices and a second chapter on probability that includes a measure-theoretic exposition. In the chapter on multivariate analysis, instead of deriving features of the multivariate normal distribution through its probability density function, Rao uses the simple characterization that every linear function of the variables has a univariate normal distribution. This book did not lend itself as well to being a course textbook as those by Mood and by Hogg and Craig, because of the high technical level, terse style, and difficult exercises. A second edition was published in 1973. Also of somewhat higher level and with strong emphasis on probability and stochastic processes was the 3rd edition of *Probability Theory and Mathematical Statistics* by Fisz (1963), which was translated from Polish to English by Robert Bartoszynski.

7 | THE QUARTER CENTURY AFTER WORLD WAR 2: SPECIALIZED BOOKS ON STATISTICAL THEORY AND METHODS

The quarter century following World War 2 also saw an explosion of books that dealt with particular aspects of statistical theory and methods. This section summarizes some books that had a major impact in the teaching of statistical science in emerging graduate programs, in the context of topics to which those programs typically devoted entire courses. Nearly all of the textbooks mentioned have the classical rather than Bayesian statistical orientation. Section 8 discusses textbooks that present the Bayesian approach.

7.1 | Experimental design and sampling methods

Following Fisher's 1935 *Experimental Design*, a useful book on this topic for statisticians and scientists was *Experimental Designs* by Cochran and Cox (1950). Three chapters present the philosophy of statistics in experimentation and basic methods of analysis. Then, 10 chapters introduce specific designs, including completely random designs, randomized blocks, Latin squares, factorial designs, split-plot designs, and various incomplete block designs. The discussion includes methods for analyzing resulting data, often including results from actual experiments. Gertrude Cox had a highly distinguished career, being a pioneer in many ways when women did not have much visibility in academia. She helped to form the Statistics Laboratory at Iowa State, was the first head of the Department of Experimental Statistics at North Carolina State and then the Institute of Statistics, founding editor of *Biometrics*, the first head of Statistics Research at the Research Triangle Institute, and a President of the American Statistical Association and of the Biometric Society. Soon after the Cochran and Cox publication, Oscar Kempthorne published *The Design and Analysis of Experiments* (Kempthorne, 1952), which presents a general statistical theory and instructions for constructing and analyzing more complex designs. Cochran and Cox published a second edition in 1957, and Klaus Hinkelmann revised Kempthorne's book for a second edition in 2007.

For observational studies, for many years William Cochran's (1953) *Sampling Techniques* and its 1962 and 1977 later editions was the classic textbook about sample surveys. Assuming knowledge of the principal topics of statistical theory, Cochran presents a wide variety of sampling methods, many of which had been developed only in the previous 15 years. Following a discussion on simple random sampling, with emphasis on finite populations, the book consider sampling for proportions and percentages and the estimation of sample size, showing the variance of estimators for the various cases. The presentation on multistage sampling emphasizes stratified random sampling. Other topics include ratio and regression estimation, systematic sampling, types of sampling unit, subsampling with units of equal and unequal sizes, and double sampling. The discussion of sources of error in sample surveys includes the effects of bias, such as due to nonresponse. At the time of its publication, it complemented Yates's (1949) *Sampling Methods for Censuses and Surveys*, which is more of a practitioner's manual.

7.2 | Testing statistical hypotheses

Testing Statistical Hypotheses by Erich Lehmann (1959) provided a highly influential treatment of the Neyman–Pearson theory and its ramifications that helped many statisticians became more familiar with it. Lehmann's book contains an enormous number of results about parametric and nonparametric approaches, with numerous exercises that themselves contain much additional material. After introducing concepts and stating basic theorems of measure theory as well as conditional probability and sufficiency, the text presents uniformly most powerful tests and related confidence intervals and the sequential probability ratio test. Two chapters then discuss uniformly most powerful unbiased tests and related confidence procedures, focusing on one- and two-sided problems in exponential families. Lehmann discusses invariance and its relation to unbiasedness and presents linear hypotheses under normality and large-sample likelihood-ratio tests. The book's depth, importance, and positive reception but lack of universal acclaim is well illustrated in a long 1961 book review by John W. Pratt in *Journal of the American Statistical Association*, which states "Lehmann has given us a definitive and beautiful mathematical treatment of an important part of the leading theory of statistics. As such, it will be invaluable to statisticians of all philosophies." However, it also states "this book, by its very excellence, its thoroughness, lucidity and precision, intensifies my growing feeling that nevertheless the theory is arbitrary, be it however 'objective,' and the problems it solves, however precisely it may solve them, are not even simplified theoretical counterparts of the real problems to which it is applied." Pratt notes the lack of justification for unbiased tests and makes other criticisms later amplified in the Bayesian statistics literature. Lehmann's book does not discuss point estimation, but he covered it later in his 1983 book *Theory of Point Estimation* (Lehmann, 1983) and a second edition of that book in 1998 with George Casella. A fourth edition of the 1959 book, co-authored by Joseph P. Romano and in two volumes, appeared in 2022.

7.3 | Multivariate statistical analysis

Anderson (1958) published *An Introduction to Multivariate Statistical Analysis*, the classic text on that subject for some time. Under the assumption of an underlying multivariate normal distribution, Anderson's textbook focuses on generalizations of univariate methods, including maximum likelihood estimation of means and variances, correlation methods, principal components, likelihood-ratio tests, the analysis of variance including Hotelling's T^2 generalization of the t-statistic, and classification problems such as discriminant analysis. While an appendix summarizes matrix theory, good knowledge of this is needed to read the book. A book review in *Econometrica* in 1959 by Maurice G. Kendall (who himself had written his own monograph on the topic, *A Course in Multivariate Analysis*, in (Kendall, 1957)) states "This will be an essential book for all serious students, in which class I include more senior statisticians" but points out that there is no discussion of how the procedures stand up to use in nonnormal variation. Books on nonparametric statistics and robustness discussed below addressed this in later years. Corresponding methods for discrete data did not receive much attention in book form until *Discrete Multivariate Analysis* by Bishop et al. (1975). A third and final edition of Anderson's book was published in 2003.

7.4 | The analysis of variance

Scheffé's (1959) *The Analysis of Variance* gave an in-depth theoretical presentation of the methods developed by Fisher 40 years earlier and of later advances. Unlike Fisher's and Snedecor's books, Scheffé's book is mainly designed for the mathematical statistician, although many examples and exercises use actual data. Part I of the book begins with the basis of least squares theory and general linear hypotheses and then presents the theory of fixed-effects models with independent observations of equal variance. Topics considered include point estimation, the construction of confidence ellipsoids and tests, multiple comparisons, partitioning sum of squares for the complete two-, three-, and higher-way layouts, Latin squares, incomplete blocks, nested designs, and the analysis of covariance. The chapter on multiple comparisons includes Scheffé's own method based on the F statistic, the Tukey method based on the studentized range, and a comparison of them. Part II of the book explores the analysis of variance for random-effects models, mixed models, and randomization models incorporating permutation tests. It also discusses the effects of departures from the underlying assumptions, such as the effects on F tests of nonnormality, heterogeneity of error variance, and nonindependence of errors. Appendices include one on matrix algebra.

7.5 | Regression analysis

Soon after its publication in 1966, *Applied Regression Analysis* by Norman Draper and Harry Smith became the standard textbook for courses on this topic. Their book begins with straight-line fitting by least squares, including discussion of a lack-of-fit test. The book then discusses multiple regression, making systematic use of matrix notation. Topics include variable selection, testing a general linear hypothesis, ANOVA tables, orthogonal polynomials, weighted least squares, and bias of estimators due to incorrect model assumptions. The book includes material not previously generally available in textbooks, such as a chapter on the examination of residuals and a chapter on selecting the "best" regression equation, which considers algorithms such as backward elimination, forward selection, and stepwise regression. Although the book is mainly concerned with multiple linear regression, it also discusses transformations of variables and shows the analysis of variance and analysis of covariance as applications of multiple regression incorporating dummy variables. The authors also devote a chapter to nonlinear estimation, including Newton–Raphson and steepest descent methods and the geometry of linear and nonlinear least squares. Not included are topics such as regression with errors in the independent variables, principal components, and the effects of departures from assumptions of normality and independence of observations. An unusual feature for that time is the assumption that results of computations are available in computer output, and an appendix contains 56 pages of printouts. The book has a practical orientation, but in subsequent years other books covered theory, such as Searle's (1971) *Linear Models*. The third and final edition of Draper and Smith appeared in 1998.

7.6 | Nonparametric statistics

Shortly after World War 2 statistical science witnessed the revolution of nonparametric methods, such as the Wilcoxon and Mann–Whitney tests. Survey books on statistical science soon began to include substantial material on nonparametric methods, an example being *Statistical Inference* by Walker and Lev (1953). Two early books dealing specifically with this area were Donald A. S. Fraser's (1957) *Nonparametric Methods in Statistics*, intended as a second course for the advanced student of mathematical statistics, and Sidney Siegel's (1956) *Nonparametric Statistics for the Behavioral Sciences*, designed for research workers with limited mathematical and statistical background. Fraser's first two chapters provide a survey of the general techniques of estimation and hypothesis testing, without any special attention to nonparametrics. For hypothesis testing, invariance theory is used to reduce several nonparametric problems to ones based on ranks. Among the problems discussed are single-sample problems of fit, location, and symmetry, randomness problems, randomized blocks and other experimental designs, and tolerance regions. A final chapter deals with large-sample consistency and relative efficiency of some tests. Siegel's book first reviews advantages and disadvantages of nonparametric methods and then treats one sample, two matched samples, two independent samples, k matched samples, k independent samples, and multivariate samples. For each technique, the discussion provides the test statistic and its distribution under the null hypothesis, separately for small and large samples. However, the book presents testing hypotheses only of "no difference" and fails to consider estimation or other types of hypotheses and pays little attention to the treatment of power. N. John Castellan Jr. co-authored a second edition in 1988. I. Richard Savage (1957) surveyed the development of nonparametric statistical methods together with a long and highly critical review of Siegel's book.

Years later, several books appeared on this topic in a brief period. *Theory of Rank Tests* by Hajek and Sidak (1967) and *Nonparametric Methods in Multivariate Analysis* by Puri and Sen (1971) provided theoretical depth. For more emphasis on applications, popular books included William J. Conover's (1971) *Practical Nonparametric Statistics*, Jean D. Gibbons's, 1971 *Nonparametric Statistical Inference*, the Hollander and Wolfe (1973) *Nonparametric Statistical Methods*, and Erich Lehmann's (1975) *Nonparametrics: Statistical Methods Based on Ranks*. During this later period, methods of robust statistics and exploratory data analysis became increasingly influential, although John Tukey's seminal *Exploratory Data Analysis* did not appear until 1977.

7.7 | Time series analysis

Maurice Kendall included substantial material about analyzing time series data in his 1961 third volume of *The Advanced Theory of Statistics* with Alan Stuart, and in 1973 he published a separate book, *Time Series*. A more influential book for modeling and forecasting time series appeared in Box and Jenkins' (1970) *Time Series Analysis: Forecasting and Control*. The book provides background material on stationary time series, introducing the autocorrelation, partial autocorrelation, and spectral density functions. Building on the Wold Representation Theorem that a covariance stationary time series can be expressed as the sum of two time series, one deterministic and one stochastic, the authors show how to approximate the dependence structure in a stationary time series using an autoregressive moving average (ARMA) model. Recognizing that in practice many time series are nonstationary, Box and Jenkins then extend the methodology by carrying out successive differencing until the resulting series of differences is stationary, leading to the autoregressive integrated moving average (ARIMA) model. For this model, they provide methods for identifying an appropriate class of processes, estimating its parameters, and checking the adequacy of its fit. Once a suitable model is selected, forecasts and prediction intervals can be generated from the fitted models. Time series with a seasonal component are also covered. A second part of the book introduces models with a transfer function that expresses the relationship between the output and an observable input series. A final part of the book considers process control, the goal of maintaining a particular level for the output by appropriate manipulation of the input series.

Many other books on time series analysis appeared about the same time, including *Spectral Analysis and Its Applications* by Jenkins and Watts (1968), which used spectral techniques to obtain a solution to the problem of model identification, *Multiple Time Series* by Hannan (1970), and *The Statistical Analysis of Time Series* by Anderson (1971). The 5th edition of Box and Jenkins, with co-authors Gregory C. Reinsel and Greta M. Ljung, appeared in 2015.

7.8 | Decision theory

Ferguson's (1967) *Mathematical Statistics: A Decision-Theoretic Approach* provides a clear presentation of the game-theory results of John von Neumann and Oskar Morgenstern based on assigning a utility to every combination of a state of nature and a course of action and to Abraham Wald's exploitation of these ideas in a decision-theoretic approach to statistical science. Topics include game theory, the main theorems of decision theory (devoted to admissibility, minimaxity, and complete class theorems), invariant statistical decision problems (including location parameter estimation, scale parameter estimation, and estimation of a continuous distribution function, using invariance to reduce the class of considered rules), testing hypotheses (including confidence regions and multiple comparisons), and sequential decision problems. The book is non-Bayesian in orientation, with Bayesian or nearly Bayesian methods being justified by their good performance, rather than a means for using prior information or subjective belief in analyzing data. The book has no discussion of nonparametrics, large-sample methods, or design, and just a bit on the basics of linear hypotheses.

The book *Theory of Games and Statistical Decisions* by Blackwell and Girschick (1954) had included more material on games but was more difficult reading and less appropriate as a textbook. Also, Ferguson presented important results that appeared after that book's publication, such as Charles Stein's inadmissibility result for the sample mean estimator of a multivariate normal mean. Ferguson's book was also more accessible than Wald's (1949) monograph *Statistical Decision Functions* yet more substantial than the 1959 book *Elementary Decision Theory* by Chernoff and Moses (1959), which was written for a first course in statistics formulated as the science of decision-making under uncertainty.

DeGroot's (1970) *Optimal Statistical Decisions* took a more actively Bayesian presentation. After a survey of probability theory, the book gives an axiomatic treatment of subjective probability and utility that argues for the inevitability of the Bayesian approach. The book shows how to use conjugate prior distributions and discusses improper prior distributions and the limiting form of posterior distributions. It presents estimation, testing hypotheses, and statistical models from a Bayesian decision-making viewpoint. Also included is a thorough discussion of sequential sampling, optimal stopping, and sequential choice of experiments, much of this presented in a textbook for the first time. This book is quite a contrast in its presentation of mathematical statistics from other books of the preceding quarter century, such as those by Mood and by Hogg and Craig.

7.9 | Categorical data analysis

In the 1960s, the number of research articles dealing with categorical data analysis began to explode. Models such as logistic regression became better known following publication of the book by David Cox on *Analysis of Binary Data* (Cox, 1970), discussed in Section 9. Other books soon followed. For example, *Statistical Methods for Rates and Proportions* by Joseph L. Fleiss (1973) was a highly readable and popular book that focused on simple techniques for comparing proportions and combining evidence under various sampling methods, applied to medical examples. However, it did not discuss confidence intervals, maximum likelihood, Bayesian inference, and statistical modeling. *Discrete Multivariate Analysis* by Bishop et al. (1975) exposed many statisticians to loglinear models for multivariate categorical responses and showed their connections with logit models.

7.10 | Textbooks on statistical methods for other disciplines

As follow-ups to the Fisher and Snedecor textbooks on statistical methods geared toward scientists, numerous books presented the methods with orientation toward particular disciplines. An excellent example is Austin Bradford Hill's (1937) *Principles of Medical Statistics*, which had 12 editions between 1937 and 1991. Other popular examples in various areas, some of which had many editions, include Blalock's (1960) *Social Statistics*, Brownlee's (1960) *Statistical Theory and Methodology in Science and Engineering*, the Sokal and Rohlf's (1969) *Biometry: The Principles and Practice of Statistics in Biological Research*, and Walter A. Shewhart's (1939) *Statistical Method from the Viewpoint of Quality Control*.

7.11 | Post-1970 areas for textbooks

By 1970, Statistics or Biostatistics departments had appeared in most major universities in North America. Although statistical science had achieved a remarkable maturity, many areas were in a relatively nascent level for university courses and textbook presentations. More years were still to pass before courses and textbooks would appear on topics now commonly seen in graduate programs, such as generalized linear models, survival analysis, longitudinal data analysis, observational data and causal modeling, missing data, statistical computing, Monte Carlo methods, high-dimensional data, and machine learning.

8 | FOUNDATIONS OF STATISTICAL SCIENCE FROM A BAYESIAN PERSPECTIVE

Theory of Probability by Cambridge professor Jeffreys (1939), is a foundational book on the Bayesian statistical approach. The title is misleading, because the book also discusses estimation and significance tests within the theory of *inverse probability*, which is Jeffreys's name for the Bayesian approach to inference. The chapter on estimation introduces exponential families and his principle of noninformative priors. In his objective Bayesian approach, he states the principle for deriving noninformative prior distributions from the sampled distribution, using Fisher information. He justifies the maximum likelihood estimate as being indistinguishable for large samples from the Bayesian estimate using a uniform prior distribution. For significance testing, Jeffreys uses Bayes factors, which he had introduced in an article four years previously. The final two chapters discuss weaknesses of other approaches. R. A. Fisher respected Jeffreys as a scientist and statistician but had criticized his approach for several years before publication of this book. Joan Fisher Box (1978, p. 441) quotes Fisher (her father) as saying that Jeffreys "makes a logical mistake on the first page which invalidates all the 395 formulae in his book." A later highly influential book for the exposition of probability itself from a Bayesian perspective was Bruno De Finetti's 1970 *Teoria della Probabilità* (translated into English in 1974 and 1975), based largely on his 1937 article.

Following Jeffreys's book, Leonard Jimmie Savage's (1954) *The Foundations of Statistics* was highly influential. Starting with six axioms and using rigorous proofs, Savage introduces the personalistic tradition in probability and the expected utility theory of von Neumann and Morgenstern and then provides a personalistic discussion of frequentist methods. The final three chapters on statistical inference primarily criticize existing frequentist inferential approaches rather than expounding positive approaches taking the Bayesian approach. For instance, he argues that users of frequentist methods "endlessly pass the buck, saying in effect, 'This assertion has arisen according to a system that will seldom lead you to make false assertions, if you adopt it. As for myself, I assert nothing but the properties of the system.' "

In a later article, Savage (1961) states that he was "too deeply in the grip of frequentist tradition to do a thorough job." He credits the textbook *Probability and Statistics for Business Decisions* by Robert Schlaifer (1959) as the first to be written "entirely and wholeheartedly from the Bayesian point of view." Although Schlaifer's detailed 742-page book does not have the theoretical depth of other post World War 2 books described in Sections 5 and 6, it was revolutionary in its Bayesian decision-making approach for handling real-world problems. Schlaifer had decided that the Pearson/Fisher/Neyman synthesis was inadequate for business applications, which he argued should have decision chosen according to the highest expected utility when one assigns values to consequences and probabilities to events. After discussing the meaning of probability and defining expected value and utility, the book presents methods, including the Monte Carlo method when a mathematical analysis is not possible. Schlaifer shows how to revise probabilities in the light of new information and considers decisions in the context of data obtained sequentially. The final part of the book discusses and criticizes the frequentist approach to statistical inference. Schlaifer argues that "the long-run frequency with which a certain method of making statements would produce incorrect statements is of no real interest to anyone" and that a statistician should be able to take account of other sources of information besides the particular sample of data. Two years later, Schlaifer co-authored *Applied Statistical Decision Theory* with Howard Raiffa (Raiffa & Schlaifer, 1961). Their book provides a more detailed and systematic Bayesian treatment that uses conjugate prior and posterior distributions to simplify analyses computationally. After introducing utility functions and Bayes' theorem, Raiffa and Schlaifer apply this approach to estimating binomial and Poisson parameters and normal mean and variance parameters and regression parameters, with and without nuisance parameters being known. They also discuss optimal sample size, stopping rules in collecting data, comparing two or multiple means, and selecting the best of several processes.

Later Statistics books having a Bayesian perspective include one by Good (1965) on estimating probabilities, a two-volume work by Dennis Lindley (1965), introductory-level textbooks by Blackwell (1969) and Schmitt (1969), and a book by Box and Tiao (1973) that emphasizes various models for a normally distributed response, mainly using noninformative prior distributions. DeGroot (1975) wrote a mathematical statistics book with a Bayesian perspective that could better serve as a course textbook. It was later revised by Mark Schervish, with a 4th edition in 2018.

9 | BOOK CONTRIBUTIONS BY DAVID R. COX

Although best known for his research articles, especially his 1972 paper that introduced the proportional hazards model for censored survival data, David Roxbee Cox was a prolific author of books. His 20 books cover an impressively broad variety of subject-matter areas. The books are concise and nearly all relatively short, typically about 200 pages or less. In an interview by Reid (1994), Cox stated

> "I find it much easier to understand something that is clearly put with a minimum number of words, when you know you've got to look at each word and think what it means. …I would in principle want to claim maximum conciseness is also maximum clarity. There's also the psychological point. I find the notion of trying to explain a certain moderately advanced subject in a couple of hundred pages, the essence of it, much more appealing than the 800-page encyclopedia on something."

Cox also mentioned that he found book-authorship to be a difficult task requiring much time, and that he had had the good fortune to have collaborations with a succession of former PhD students and friends. In fact, of the 20 books summarized in this section, only four were sole-authored.

While working on his PhD, which he received from the University of Leeds in 1949, and before moving to the Statistical Laboratory at the University of Cambridge, Cox was employed by the Wool Industries Research Association in Leeds from 1946 to 1950. In 1949 he co-authored *An Outline of Statistical Methods for Use in the Textile Industry* with Alan Brearley (Cox & Brearley, 1949). This short (74 page) publication had six editions between 1949 and 1960.

Cox (1958) published *Planning of Experiments* in 1958. As in his other books, he focuses on basic concepts rather than technical mathematical and computational details, so the book is accessible to experimenters as well as statisticians. Following introductory chapters on assumptions underlying the designs and their analyses, such as additivity and constancy of treatment effects, Cox discusses designs for error reduction, such as randomized blocks, Latin squares, and covariance adjustments. A chapter then discusses the nature of, and reasons for, randomization. Subsequent topics include factorial designs, the choice of sample size, sequential methods in comparative experimentation, incomplete designs, fractional replication, confounding, and cross-over designs. Cox does not attempt to explain statistical analyses for the designs, leaving him free to concentrate on the design aspects. For instance, he mentions analysis of variance but shows no examples of it. Drawing on his experience as a statistician working at the Royal Aircraft Establishment, Wool Industries Research Association, and the Statistics Laboratory at the University of Cambridge, Cox uses examples from a wide variety of applications. Cox wrote the book at a young age, shortly after he had taken a position as Reader at Birkbeck College. It received outstanding reviews in statistics journals by highly-esteemed statisticians such as Peter Armitage, William Cochran, and F. N. David. In his review in *Biometrika* in 1959, Colin Mallows concludes "This is a book about real statistics. …It should be required reading for all students of statistics."

Cox's 1961 monograph *Queues* was co-authored with Walter L. Smith, whose Cambridge PhD in 1953 had been co-directed by Cox and Henry Daniels (Cox & Smith, 1961). Queues for service arise in many fields, and this monograph concisely introduces the theory of queueing systems and mathematical techniques useful for studying them. Apart from some familiarity with generating functions and Laplace transforms, the book assumes no advanced mathematical knowledge. Following an introductory chapter, Cox and Smith focus on a single server with random arrivals for exponential and general service-time distributions, with special emphasis on equilibrium conditions and the distribution of busy periods. The monograph also considers nonequilibrium theory, queues with many servers, queues with priorities, and the use of Monte Carlo methods to simulate queueing behavior.

In 1962 Cox published a companion monograph on *Renewal Theory* (Cox, 1962), apparently the first book devoted exclusively to this topic, although Walter L. Smith had written a long discussion paper on the topic in *Journal of the Royal Statistical Society, Series B* in 1958. For a random variable X_1 and independent, identically distributed nonnegative random variables $\{X_2,\ldots,X_n\}$ that represent lengths of time between certain recurrent events (renewals), $R_n = \sum_{i=1}^{n} X_i$ is the time at which renewal n takes place. Renewal theory is concerned with the distribution of the number of renewals that occur within a given interval of time, the moments of the number of renewals, the length of time at any given instant until the next renewal or since the last renewal, and their asymptotic properties. Assuming some familiarity with probability, the book presents fundamental models for the renewal process in continuous time in a style intended for students and workers in operations research and allied disciplines. Special situations considered include superpositions of renewal processes in which several independent renewal processes are considered simultaneously and renewals are recorded whenever a renewal takes place in any one of the processes, alternating renewal processes of form $\{X_1, Y_1, X_2, Y_2,\ldots\}$ in which a component of one type is always replaced on failure by a component of a second type and in which $\{X_i\}$ and $\{Y_i\}$ have separate distributions, and probabilistic models of failure with strategies for replacement of an item due to its wear.

In 1965 Cox co-authored *The Theory of Stochastic Processes* with Hilton D. Miller, at that time a Reader at Birkbeck College and later a highly-accomplished artist (https://hiltonmiller.com; Cox & Miller, 1965). Their book, which assumes knowledge of elementary probability theory, standard results in matrix algebra, and Laplace and Fourier transforms, illustrates topics with examples from several scientific fields. Cox and Miller discuss Markov chains, illustrating with random walks and branching processes. They then present Markov processes with discrete states in continuous time, such as the Poisson process and birth and death processes. The following material on Markov processes with continuous state space in continuous time deals almost entirely with the Wiener process and its diffusion equations, first passage times, behavior with absorbing or reflecting barriers, and transformations. Cox and Miller also cover point processes and non-Markovian processes using techniques from queueing theory.

In 1966 Cox published *The Statistical Analysis of Series of Events*, written with Peter A. W. Lewis, a PhD student of Cox's at Birkbeck College while Lewis was employed by the IBM Research Laboratories in San Jose, California (Cox & Lewis, 1966). The book presents methods for observations in the form of a series of point events that occur in a continuum, typically in time but also possibly in space. Examples are the times of breakdown in a piece of equipment, times of accidents, and arrival times of phone calls. The simplest model for such data is the Poisson process. The book builds on discussion papers in *Journal of the Royal Statistical Society, Series B* by

Lewis in 1964 and by Cox in 1955, which presented generalizations of the Poisson process having varying rates, including a doubly stochastic extension now known as the Cox process. After introducing the Poisson process, Cox and Lewis show how to analyze whether the point events structure is changing with time. The remainder of the monograph is concerned with stationary processes. Two chapters deal with theoretical aspects and present methods that involve only the means, variances and correlations, using spectral analysis and serial correlation techniques. Two chapters discuss renewal processes, such as generalizing the Poisson process to allow the independent intervals between successive events to have other than an exponential distribution. Cox and Lewis also discuss distribution-free goodness-of-fit tests, a process that permits autocorrelation in renewal intervals, a branching renewal process with application to computer failure data, a two-state semi-Markov process, estimation of parameters of component processes from observations of a pooled process, and methods for comparing two Poisson processes.

Although books were available by the 1960s on regression analysis, such as Draper and Smith (1966), none focused on logistic regression modeling. Thus, Cox's (1970) monograph *Analysis of Binary Data*, which presents a unified approach to such models for binary response variables, had a major impact. Although proposed 25 years previously by Joseph Berkson and developed further in a discussion paper by Cox in *Journal of the Royal Statistical Society, Series B* in 1958, logistic regression was not yet commonly used. Partly this reflected the lack of familiarity with it by most statisticians, but also the computational difficulty of model fitting, compared with ordinary linear regression. Cox shows how to obtain large-sample efficient estimates using weighted least squares and how to exploit the model's simple sufficient statistics to perform an exact, conditional analysis for small samples. He notes that the logistic approach is the only one for binary data to have the same simple sufficient statistics as the ordinary linear model. Cox also discusses the graphical analysis of residuals, which he had treated more generally in a highly cited 1968 article in *Journal of the Royal Statistical Society, Series B* with E. Joyce Snell, a colleague at Imperial College. Cox prepared a second edition in 1989 with Snell, increasing the emphasis on maximum likelihood methods for model fitting.

Following the 1940s-era mathematical statistics textbooks discussed in Section 5 and their successors discussed in Section 6, in 1974 David Cox and David Hinkley (whose PhD thesis had been supervised by Cox at Imperial College) presented their own view of the foundations of the subject in *Theoretical Statistics* (Cox & Hinkley, 1974). Sacrificing mathematical rigor and avoiding a theorem-proof style in order to emphasize general concepts, the authors begin with an overview of basic philosophical ideas behind statistical procedures, introducing likelihood and sufficiency as central concepts. They then cover hypothesis testing for simple and composite null hypotheses, with emphasis on the Neyman–Pearson lemma, efficient scores, similarity, conditioning, and invariance. The book carefully distinguishes between Fisher's significance testing and the Neyman–Pearson approach. Subsequent chapters discuss distribution-free and randomization tests, interval estimation based on tests, point estimation for minimizing variance or mean squared error with focus on maximum likelihood, asymptotic theory, Bayesian methods, and decision theory. A review by S. D. Silvey (who published his own book on the topic, *Statistical Inference*, in Silvey (1970) in *Journal of the Royal Statistical Society, Series A*); begins with "With all due respect to his co-author Dr Hinkley, it is difficult not to regard this book as Cox's *magnum opus*, and as such it makes compulsive reading for the professional statistician. …Indeed it is difficult to think of any other book of comparable size in which so many concepts are discussed, and this is partly what makes it essential reading for the professional statistician." The book's exercises are very challenging, and to accompany the book, in 1978 Cox and Hinkley published *Problems and Solutions in Theoretical Statistics* (Cox & Hinkley, 1978).

The 1965 Cox and Miller book on stochastic processes ends with a chapter on point processes, and the 1966 Cox and Lewis book considers series of point events. This material was extended further in 1980 with *Point Processes*, co-authored with Valerie Isham, whose PhD thesis Cox supervised at Imperial College (Cox & Isham, 1980). Assuming knowledge of basic probability theory and some background with Markov processes, the book emphasizes methods useful in a variety of applications. An introductory chapter shows examples such as Poisson processes and renewal processes. The authors then present a theoretical framework, followed by details about specific models such as those introduced in the first chapter, operations on point processes, multivariate point processes, and spatial processes.

Cox and Snell (1981) published *Applied Statistics, Principles and Examples*. The first 50 pages discuss the nature, objectives, and strategy of statistical analysis, assuming that the reader has some experience with data analysis and knowledge of statistical methods, including least squares, analysis of variance, and maximum likelihood. The authors distinguish among five types of investigation: controlled experiments, pure observational studies, sample surveys, controlled prospective studies, and controlled retrospective studies. The rest of the book shows examples, each one presenting the questions to be answered, strategy adopted, and a methodological analysis that first inspects the data and constructs simple descriptive statistics. Methods include normal linear models, analysis of variance, and logistic and loglinear models.

Cox's landmark 1972 *Journal of the Royal Statistical Society, Series B* paper "Regression models and life-tables" introduced the proportional hazards model for investigating the relationship between survival times of patients and explanatory variables with censored data. The 1984 book *Analysis of Survival Data* by Cox and David Oakes, a former PhD student of Cox's at Imperial College, presents the primary methods for analyzing nonnegative random variables with possible right censoring (Cox & Oakes, 1984). Directed toward applied statisticians, applications include modeling life expectancy using explanatory variables such as treatment and individual histories. Examples include the Stanford heart transplant program and industrial life testing. Following an introductory chapter about failure times and censoring, the book discusses distributions for failure times, introduces the hazard function, and presents a parametric analysis for a single sample. This material includes likelihood theory, a standardized mortality ratio that is the maximum likelihood estimate of relative risk for exponentially distributed failure times, the proportional hazards family, and the Weibull distribution. The book then presents nonparametric methods, including the product-limit estimator and the accelerated life model. The heart of the book includes explanatory variables in the models, with fully parametric approaches and Cox's proportional hazards model. The discussion of time-dependent covariates introduces partial likelihood and compares efficiency with fully parametric models. Other topics include multiple failure times with independent competing risks, bivariate survivor functions, and estimation using the EM algorithm.

Cox's books discussed so far, except for *Theoretical Statistics*, were designed for an audience that included both applied statisticians and statistically-literate scientists. His next two books, co-authored with Danish statistician Ole Barndorff-Nielsen, were instead directed toward those having a strong background in probability and statistical theory. Barndorff-Nielsen and Cox's (1989) *Asymptotic Techniques for Use in Statistics* deals with methods for handling asymptotic calculations in statistics. The most basic of these, using local linearization (the delta method) and approximate normality, have a long history. The methods give simple approximate answers to parametric distributional problems where an exact solution is computationally intractable. With higher-order expansions, the accuracy of simple approximations can be assessed and improved. Following an introduction to properties of sums of random variables and basic ideas of convergence of random variables, the book contains an extensive treatment of asymptotic expansions derived from integrals and series, such as Edgeworth series and multivariate cumulant expansions, large deviation theory, and the method of tilting and its uses for univariate and multivariate expansions of smooth functions of vector means of independent random variables. Barndorff-Nielsen and Cox's (1994) *Inference and Asymptotics* discusses more recent developments in asymptotic parametric inference from a likelihood-based perspective, including profile, adjusted, conditional, and penalized likelihoods. After providing a review of key concepts such as likelihood, sufficiency, conditionality, ancillarity, parameter orthogonality, and full and curved exponential families, the book presents first-order and higher-order asymptotic theory. An especially fruitful approach applies the p^* formula that Barndorff-Nielsen had proposed in a 1983 article in *Biometrika*. It approximates the conditional distribution of the maximum likelihood estimator, given an ancillary statistic, using the normed log-likelihood and the determinant of the observed information matrix.

Cox's 1996 book *Multivariate Dependencies: Models, Analysis and Interpretation* with Nanny Wermuth, who at the time was a professor at the University of Mainz, discusses graphical representations of dependencies and independencies among variables, with emphasis on large observational studies in the social sciences (Cox & Wermuth, 1996). The graphical representation uses nodes to represent variables, edges with directed arrows to represent relations between explanatory variables and response variables, and lines without arrowheads to join variables treated on an equal footing. Intermediate variables can serve as responses to some variables and as explanatory to others. Such representations are useful for the study of implications of a given model regarding conditional independencies as well as for concepts of causality. The book assumes knowledge of regression analysis, but most of the book keeps the mathematical level low to make it accessible to researchers who use statistical methods.

Cox wrote *The Theory of Design of Experiments* (Cox & Reid, 2000) with University of Toronto professor Nancy M. Reid, who had done postdoctoral research with Cox at Imperial College. Written with a relatively high technical level compared with Cox's (1958) book on this topic, their book stresses that successful experimental designs adapt general principles to the constraints of particular applications. Following an introduction, the authors discuss the avoidance of bias through randomization and retrospective adjustment. They then explain control of haphazard variation, with emphasis on blocking, discussing matched pairs, the randomized block design, Latin squares, incomplete block designs, and cross-over designs. This is followed by material on factorial designs, confounding, fractional factorials, split-plot and Taguchi designs, and response surface methods. A chapter on optimal designs includes nonlinear designs and Bayesian designs. The book concludes with discussion of sample size determination, adaptive designs, and spatial designs. Like Cox's 1958 book, the emphasis is on design rather than analysis, although appendices summarize some statistical analyses for linear models and computational issues using S-Plus.

In 2000 Cox also edited *Complex Stochastic Systems*, with Ole Barndorff-Nielsen and Claudia Klüppelberg (Barndorff-Nielsen et al., 2000). In it, leading researchers addressed various statistical aspects of the field, such as Markov chain Monte Carlo, causal inference from graphical models, and hidden Markov models, illustrated by applications.

The 2003 book *Components of Variance*, with Patricia J. Solomon, formerly a PhD student of Cox's at Imperial College, was written primarily for statisticians (Cox & Solomon, 2003). They first present key models and concepts, explaining the distinction between nesting and cross-classification and between fixed and random effects. They discuss the balanced case of one-way analysis of variance with random effects in considerable detail and then cover more general balanced contexts. Their discussion of unbalanced situations moves from the one-way classification for the linear model to logistic regression modeling, including brief discussion of REML and meta-analysis. In addition to Poisson and binomial models, Cox and Solomon discuss survival data and extensions to generalized linear mixed models. The final chapter considers ways of assessing models, including the study of outliers and nonparametric estimation of the distributional form of the underlying random variables. The varied examples include ones from genetic data analysis, clinical trial design, longitudinal data analysis, industrial design, and meta-analysis.

Cox (2006) wrote his first sole-authored book since 1970, *Principles of Statistical Inference*. Assuming some previous knowledge of statistics but keeping the mathematics at an elementary level without using derivations or proofs, Cox presents the key concepts and principles. He compares the frequentist and Bayesian approaches to statistical inference in an undogmatic style that discusses advantages and disadvantages of each. Following preliminaries, Cox presents a Fisherian frequentist approach of finding the likelihood function for an exponential family distribution, the sufficient statistic, and a pivot based on the sufficient statistic for the parameter of interest that is used to find *p*-values and confidence intervals. An extended version of this approach involves conditioning on an ancillary statistic, with asymptotics using the maximum likelihood estimate and the observed Fisher information as approximate sufficient statistics. Cox then discusses significance testing, including the relation with interval estimation and Bayesian testing, but indicates a preference for supplying *P*-values and confidence intervals without the Neyman–Pearson approach of focusing on decisions and power maximization. Chapters on asymptotic theory and maximum likelihood discuss nuisance parameters, profile likelihood, higher-order asymptotics, multimodal likelihoods, singular information matrices, and modified likelihoods. The book also considers nonlikelihood-based methods and randomization-based analysis. Two appendices give a brief historical overview and a personal assessment of the merits of different ideas, revealing Cox's preference for the frequentist approach.

Cox and Donnelly (2011) co-wrote a companion book, *Principles of Applied Statistics*. Donnelly was then a professor at Imperial College, but Cox had been her statistical mentor after she moved to Oxford in 1995. Cox and Donnelly assume some knowledge of standard statistical methods but de-emphasize mathematical theory. After introducing general concepts, the authors discuss the "analysis pipeline," with careful attention to measurement, dimension analysis, data collection, model formulation, and interpretation. The detailed discussion of study design covers units of analysis, principles of measurement, choice of scales, latent variables, avoidance of systematic error, and experimental and prospective and retrospective observational studies. Preliminary analyses include data auditing and screening, graphical, and tabular analyses. With a parametric rather than nonparametric focus, Cox and Donnelly discuss the formulation and choice of a probability model, confidence intervals and significance tests for formal inference, multiple testing, and interpretation including issues of statistical causality and publication bias.

Cox's final book is the 2014 *Case-Control Studies*, co-authored with Ruth H. Keogh of the London School of Hygiene and Tropical Medicine, whom Cox had supervised for her Masters and PhD degrees at Oxford (Keogh & Cox, 2014). The case-control approach is extensively used in epidemiology to study disease incidence. For example, Austin Bradford Hill and Richard Doll employed it in 1950 to compare smoking histories of patients hospitalized with lung cancer and without lung cancer to investigate the possible association between lung cancer and cigarette smoking. Assuming background in statistical theory and methods, Cox and Keogh emphasize ways to construct likelihood functions using relevant conditional probabilities for estimating parameters such as odds ratios. Following an introduction, including methods for controlling confounding, temporal aspects, and biases arising under retrospective ascertainment, they present the key characteristics of unmatched and matched case-control studies. They begin with the simplest situation of a binary exposure and binary outcome and then present a general formulation of logistic regression for unmatched case-control studies under a population model, a sampling model, and an inverse model for the conditional probability of risk factors given the outcome. A fundamental result is the equivalence of the maximum likelihood estimate of the odds ratio under prospective and retrospective sampling. Cox and Keogh also discuss non-logistic models, studies with more than two outcomes or more than one control group, nested case-control studies, case-subcohort studies that are appropriate for sampling within cohorts while accounting for time, and misclassification and measurement error. They conclude by presenting ways to synthesize results of studies, such as by using a random effects framework. An appendix provides theoretical background on maximum likelihood and pseudo likelihood.

Generally, Cox's books were not intended to be course textbooks. In discussing writing *Planning of Experiments* in Reid (1994), Cox stated

> "It isn't a textbook, you see. In fact in a certain way none of the books I've written are textbooks. None of the books I've written are other than extremely indirectly based on lectures, they're not really meant to teach courses from. They are attempts to write down a subject that I've thought about for a while, as it seems to me, or whoever I've been working with, and I've been very very fortunate in the people I've worked with. Maybe it would have been better if they had been textbooks, but they aren't."

He explains his reasoning for writing books in this manner:

> "I would not normally, even for a moment, consider using a textbook in a course of lectures. I would refer the students to several different books, and see the role of books as backup for the teacher of a course, the opportunity to choose what he or she thought was important. Not to set out an exact prescription for somebody to follow."

As also emphasized in the summaries in this section, Cox's books were usually intended for a relatively broad scientific audience, assuming merely some statistical and mathematical background. As a consequence, Cox undoubtedly had much more influence among applied statisticians and scientists than if he had written only technical research articles. Finally, after writing a book on a particular subject, Cox seemed to prefer to move on to other topics, often quite different, for subsequent ones. In fact, only one of his books (*Analysis of Binary Data*); Cox's (1970) has had a second edition.

10 | THE FUTURE OF TEXTBOOKS ON STATISTICAL FOUNDATIONS IN THE DATA SCIENCE ERA

Except for the previous section on books of Sir David Cox, this article has focused on contributions in the first half of the 20th century by Yule, Fisher, Snedecor, Kendall, Wilks, and Cramér and then books in the following quarter century by prominent statisticians that helped statisticians and scientists to more fully recognize foundations for this rapidly developing field. It is an ever-increasing challenge to write a textbook about the foundations of statistical science, because of the growth of the field, with continual introduction of new methods and new types of data and a broader scope that includes data preparation and visualization. So, what form should such a textbook take, now and in the future? Since the 1900–1970 time period on which this article has focused, statistical science has centered increasingly on computational methods and applications and less on mathematical formulations and proofs. Reflecting the revolution instigated by Tukey (1977) in the data analysis expansion of the field and by Breiman (2001) in formulating an algorithmic alternative to a modeling culture, mathematical statistics is likely to increasingly be viewed merely as a small part of the broad field of data science. It now seems artificial for a textbook to place a boundary between mathematical statistics and methods of data analysis or to have purely a frequentist or a Bayesian focus. For example, in a book that I recently co-authored (Agresti & Kateri, 2022) that is designed to teach foundational concepts and results to undergraduate students who plan to be data scientists, the titles of most chapters did not differ much from those in a mathematical statistics book such as Alexander Mood's seventy years ago, but we illustrated key frequentist and Bayesian results using simulations, a variety of apps such as Bernhard Klingenberg's outstanding apps at https://www.artofstat.com, and examples employing R and Python.

In the future, books on the foundations of statistical science should and probably will include more material on aspects of machine learning and the analysis of high-dimensional data, such as now presented in books such as Efron and Hastie (2016), Izenman (2008), and Wainwright (2019). The contents of these books differ dramatically from the main topics introduced by Cramér, Kendall, and Wilks that set the tone for statistical science for more than half a century. In the companion article to this one that focused on statistical theory textbooks, Agresti (2021) asked several renowned professors including Sir David Cox to contribute their thoughts about the presentation of the foundations of statistical science in the modern data science era or about the historical evolution of textbooks on the subject. In his comments, Bradley Efron stated "Cox and Hinkley's 1974 book *Theoretical Statistics* shows that it is possible to navigate deep statistical waters without getting stuck in the Sargasso Sea of asymptotics. A contemporary version that took on 21st Century topics would be most welcome." Even with high-dimensional data, classical theory can provide a starting point for analysis. Xiao-Li Meng pointed out that if data science is to be a sound discipline in the future, it is especially relevant to provide statistical foundations for learning algorithms that form its core. Although forms of data are increasingly diverse, basic principles of long-existing topics such as the design of experiments will always be relevant in assessing the quality of the data and consequent description and inference. Complementary to these discussions, many articles have now been published about future directions for the field in the data science era, such as Arjas (2011) and Donoho (2017).

Teaching also needs to reflect changes in the scope of statistical science. Browsing the Internet, one finds that most universities still have a course called "mathematical statistics" or something similar. However, a course in statistical theory, at the undergraduate or graduate level, should now take into account the breadth of the field, removing the boundary between mathematical and computational ways of dealing with variation and uncertainty and between frequentist and Bayesian approaches. Teaching in introductory service courses should highlight principles and methods without getting bogged down in computational details, a trend that has become increasingly popular in recent years.

Finally, although this article has focused on the evolution of textbooks between 1900 and 1970, naturally a strong understanding of statistical science and the recent data science generalizations requires reading and studying foundational articles in research journals. In his comments in Agresti (2021), David Cox stated "The trouble with restricting to textbooks is a distorted picture of the history, I think. In my view, subjective judgement of course, the three most important publications up to 1940 are Fisher's 1922 paper 'On the mathematical foundations of theoretical statistics' (*Philosophical Transactions of the Royal Society A*), his 1925 paper 'Theory of statistical estimation' (*Proceedings of the Cambridge Philosophical Society*), and his 1926 paper 'The arrangement of field experiments' (*Journal of the Ministry of Agriculture of Great Britain*) which led to his 1935 book *The Design of Experiments*." As we statisticians view our past and think about our future, we should pay attention to the wisdom that Sir David Cox expressed in many ways over the course of his career. I myself was lucky to get to know him during a 1981–1982 sabbatical year spent at Imperial College, and I am pleased to dedicate this article to his memory.

ACKNOWLEDGMENTS

Thanks to two referees, Nicholas J. Cox, Jay Devore, David Hitchcock, and Greta Ljung for helpful comments and corrections. Thanks also to Sir David Cox and Stephen Stigler for comments and suggestions about related material in the companion article (Agresti, 2021).

ORCID

Alan Agresti https://orcid.org/0000-0003-2012-3166

REFERENCES

Agresti, A. (2021). The foundations of statistical science: A history of textbook presentations. *Brazilian Journal of Probability and Statistics*, *35*, 657–698.

Agresti, A., & Kateri, M. (2022). *Foundations of statistics for data scientists, with R and python*. CRC Press.

Agresti, A., & Meng, X.-L. (2013). *Strength in numbers: The rising of academic statistics departments in the U.S.* Springer.

Aitken, A. C. (1939). *Statistical mathematics*. Oliver and Boyd.

Anderson, T. W. (1958). *An introduction to multivariate statistical analysis*. Wiley.

Anderson, T. W. (1971). *The statistical analysis of time series*. Wiley.

Arjas, E. (2011). On future directions in statistical methodologies – Some speculations. *Scandinavian Journal of Statistics*, *38*, 185–194.

Barndorff-Nielsen, O. E., & Cox, D. R. (1989). *Asymptotic techniques for use in statistics*. Chapman & Hall/CRC.

Barndorff-Nielsen, O. E., & Cox, D. R. (1994). *Inference and asymptotics*. Chapman & Hall/CRC.

Barndorff-Nielsen, O. E., Cox, D. R., & Klüppelberg, C. (2000). *Complex stochastic systems*. Chapman & Hall/CRC.

Bishop, Y. M. M., Fienberg, S. E., & Holland, P. W. (1975). *Discrete multivariate analysis*. MIT Press.

Blackwell, D. (1969). *Basic statistics*. McGraw-Hill.

Blackwell, D., & Girschick, M. A. (1954). *Theory of games and statistical decisions*. Wiley.

Blalock, H. M. (1960). *Social statistics*. McGraw-Hill.

Bowley, A. L. (1901). *Elements of statistics*. P. S. King & Son.

Box, G. E. P., & Jenkins, G. M. (1970). *Time series analysis: Forecasting and control*. Holden-Day.

Box, G. E. P., & Tiao, G. C. (1973). *Bayesian inference in statistical analysis*. Addison-Wesley.

Box, J. F. (1978). *R. A. Fisher: The life of a scientist*. Wiley.

Breiman, L. (2001). Statistical modeling: The two cultures. *Statistical Science*, *16*, 199–231.

Brownlee, K. A. (1960). *Statistical theory and methodology in science and engineering*. Wiley.

Brunk, H. D. (1960). *An introduction to mathematical statistics*. Ginn and Company.

Chernoff, H., & Moses, L. E. (1959). *Elementary decision theory*. Wiley.

Cochran, W. G. (1953). *Sampling techniques*. Wiley.

Cochran, W. G., & Cox, G. (1950). *Experimental designs*. Wiley.

Cox, D. R. (1958). *Planning of experiments*. Wiley.

Cox, D. R. (1962). *Renewal theory*. Methuen.

Cox, D. R. (1970). *Analysis of binary data* (2nd ed.). Methuen edition 1989 with Snell E. Joyce, Chapman & Hall/CRC.

Cox, D. R. (2006). *Principles of statistical inference*. Cambridge University Press.

Cox, D. R., & Brearley, A. (1949). *An outline of statistical methods for use in the textile industry*. Wool Industries Research Association.

Cox, D. R., & Donnelly, C. A. (2011). *Principles of applied statistics*. Cambridge University Press.
Cox, D. R., & Hinkley, D. V. (1974). *Theoretical statistics*. Chapman & Hall.
Cox, D. R., & Hinkley, D. V. (1978). *Problems and solutions in theoretical statistics*. Chapman & Hall.
Cox, D. R., & Isham, V. (1980). *Point processes*. Chapman & Hall/CRC.
Cox, D. R., & Lewis, P. A. W. (1966). *The statistical analysis of series of events*. Chapman & Hall.
Cox, D. R., & Miller, H. D. (1965). *The theory of stochastic processes*. Wiley.
Cox, D. R., & Oakes, D. (1984). *Analysis of survival data*. Chapman & Hall/CRC.
Cox, D. R., & Reid, N. M. (2000). *The theory of design of experiments*. Chapman & Hall/CRC.
Cox, D. R., & Smith, W. L. (1961). *Queues*. Methuen.
Cox, D. R., & Snell, J. (1981). *Applied statistics, principles and examples*. Chapman & Hall/CRC.
Cox, D. R., & Solomon, P. J. (2003). *Components of variance*. Chapman & Hall/CRC.
Cox, D. R., & Wermuth, N. (1996). *Multivariate dependencies: Models, analysis and interpretation*. Chapman & Hall.
Cramér, H. (1946). *Mathematical methods of statistics*. Princeton University Press.
de Finetti, B. (1974, 1975). *Theory of probability (two volumes translated by A. Machi and A. F. M. Smith)*. Wiley.
DeGroot, M. (1970). *Optimal statistical decisions*. McGraw-Hill.
DeGroot, M. (1975). *Probability and statistics*. Addison Wesley.
Donoho, D. (2017). 50 years of data science. *Journal of Computational and Graphical Statistics*, *26*, 745–766.
Draper, N. R., & Smith, H. (1966). *Applied regression analysis*. Wiley.
Efron, B., & Hastie, T. (2016). *Computer age statistical inference*. Cambridge University Press.
Ferguson, T. (1967). *Mathematical statistics: A decision-theoretic approach*. Academic Press.
Fisher, R. A. (1925). *Statistical methods for research workers*. Oliver & Boyd.
Fisz, M. (1963). *Probability theory and mathematical statistics*. Wiley.
Fleiss, J. L. (1973). *Statistical methods for rates and proportions*. Wiley.
Fraser, D. A. S. (1957). *Nonparametric methods in statistics*. Wiley.
Fraser, D. A. S. (1958). *Statistics: An introduction*. Wiley.
Freund, J. E. (1962). *Mathematical statistics*. Prentice-Hall.
Gibbons, J. D. (1971). *Nonparametric statistical inference*. McGraw Hill.
Good, I. J. (1965). *The estimation of probabilities: An essay on modern Bayesian methods*. MIT Press.
Hajek, J., & Sidak, Z. (1967). *Theory of rank tests*. Academic Press.
Hald, A. (1952). *Statistical theory with engineering applications*. Wiley.
Hannan, E. J. (1970). *Multiple time series*. Wiley.
Hill, A. B. (1937). *Principles of medical statistics*. Lancet.
Hogg, R. V., & Craig, A. T. (1959). *Introduction to mathematical statistics*. Macmillan.
Hollander, M., & Wolfe, D. A. (1973). *Nonparametric statistical methods*. Wiley.
Hotelling, H. (1940). The teaching of statistics. *Annals of Mathematical Statistics*, *11*, 457–470.
Hotelling, H. (1951). The impact of R. A. Fisher on statistics. *Journal of the American Statistical Association*, *46*, 35–46.
Izenman, A. J. (2008). *Modern multivariate statistical techniques: Regression, classification, and manifold learning*. Springer.
Jeffreys, H. (1939). *Theory of probability*. Clarendon Press.
Jenkins, G. M., & Watts, D. G. (1968). *Spectral analysis and its applications*. Holden Day.
Kempthorne, O. (1952). *The design and analysis of experiments*. Wiley.
Kendall, M. G. (1943). *The advanced theory of statistics* (Vol. *1*). Griffin.
Kendall, M. G. (1946). *The advanced theory of statistics* (Vol. *2*). Griffin.
Kendall, M. G. (1957). *A course in multivariate analysis*. Griffin.
Kendall, M. G., & Stuart, A. (1958). *The advanced theory of statistics Distribution theory* (Vol. *1*). Griffin.
Kendall, M. G., & Stuart, A. (1961a). *The advanced theory of statistics Inference and relationship* (Vol. *2*). Griffin.
Kendall, M. G., & Stuart, A. (1961b). *The advanced theory of statistics Design and analysis, and time series* (Vol. *3*). Griffin.
Keogh, R. H., & Cox, D. R. (2014). *Case-control studies*. Cambridge University Press.
Lehmann, E. L. (1959). *Testing statistical hypotheses*. Wiley.
Lehmann, E. L. (1975). *Nonparametrics: Statistical methods based on ranks*. Holden-Day.
Lehmann, E. L. (1983). *Theory of point estimation*. Wiley.

Lehmann, E. L. (2008). *Reminiscences of a statistician*. Springer.
Lindgren, B. W. (1962). *Statistical theory*. Macmillan.
Lindley, D. V. (1965). *Introduction to probability and statistics from a Bayesian viewpoint; volume 1 Probability, volume 2 Inference*. Cambridge University Press.
Mood, A. M. (1950). *An introduction to the theory of statistics*. McGraw-Hill.
Puri, M. L., & Sen, P. K. (1971). *Nonparametric methods in multivariate analysis*. Wiley.
Raiffa, H., & Schlaifer, R. (1961). *Applied statistical decision theory*. Harvard Business School Division of Research.
Rao, C. R. (1965). *Linear statistical inference and its applications*. Wiley.
Reid, N. (1994). A conversation with Sir David Cox. *Statistical Science*, *9*, 439–455.
Rietz, H. L. (1927). *Mathematical statistics*. Mathematical Association of America.
Savage, I. R. (1957). Nonparametric statistics. *Journal of the American Statistical Association*, *52*, 331–344.
Savage, L. J. (1954). *The foundations of statistics*. Dover.
Savage, L. J. (1961). *The foundations of statistics reconsidered*. In J. Neyman (Ed.), *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability Contributions to the theory of statistics* (Vol. *1*, pp. 575–586). University of California.
Scheffé, H. (1959). *The analysis of variance*. Wiley.
Schlaifer, R. (1959). *Probability and statistics for business decisions*. McGraw-Hill.
Schmitt, S. A. (1969). *Measuring uncertainty: An elementary introduction to Bayesian statistics*. Addison-Wesley.
Searle, S. R. (1971). *Linear models*. Wiley.
Shewhart, W. (1939). *Statistical method from the viewpoint of quality control*. U.S. Department of Agriculture.
Siegel, S. (1956). *Nonparametric statistics for the behavioral sciences*. McGraw-Hill.
Silvey, S. D. (1970). *Statistical inference*. Penguin.
Snedecor, G. W. (1937). *Statistical methods, subtitle applied to experiments in agriculture and biology*. Collegiate Press.
Sokal, R. R., & Rohlf, F. J. (1969). *Biometry: The principles and practice of statistics in biological research*. Freeman.
Tukey, J. (1977). *Exploratory data analysis*. Addison Wesley.
Wainwright, M. J. (2019). *High-dimensional statistics: A non-asymptotic viewpoint*. Cambridge University Press.
Wald, A. (1949). *Statistical decision functions*. Wiley.
Walker, H. M., & Lev, J. (1953). *Statistical inference*. Holt.
Weatherburn, C. E. (1946). *A first course in mathematical statistics*. Cambridge University Press.
West, C. J. (1918). *Introduction to mathematical statistics*. R. G. Adams and Company.
Wilks, S. S. (1947). *Mathematical statistics*. Princeton University Press.
Wilks, S. S. (1962). *Mathematical statistics*. Wiley.
Yates, F. (1949). *Sampling methods for censuses and surveys*. Charles Griffin and Co.
Yule, G. U. (1911). *An introduction to the theory of statistics*. Charles Griffin and Co.

How to cite this article: Agresti, A. (2023). A historical overview of textbook presentations of statistical science. *Scandinavian Journal of Statistics*, *50*(4), 1641–1666. https://doi.org/10.1111/sjos.12641