What Is Statistics?

Stephen E. Fienberg

Department of Statistics, Heinz College, and Machine Learning Department, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213-3890; email: fienberg@stat.cmu.edu

Annu. Rev. Stat. Appl. 2014. 1:1–9

The *Annual Review of Statistics and Its Application* is online at statistics.annualreviews.org

This article's doi: 10.1146/annurev-statistics-022513-115703

Keywords

cross-disciplinary research, data analysis, statistics as a discipline, statistical thinking

Abstract

One might think that there is a simple answer to the question posed in the title of the form "Statistics is. . . ." Sadly, there is not, although many contemporary statistical authors have attempted to answer the question. This article captures the essence of some of these efforts, setting them in their historical contexts. In the process, we focus on the cross-disciplinary nature of much modern statistical research. This discussion serves as a backdrop to the the aims of the *Annual Review of Statistics and its Application*, which begins publication with the present volume.

1. INTRODUCTION

The title of this essay is not original. Wilson (1927), Bartholomew (1995), and Brown & Kass (2009) are but three examples of those who have used the same title for articles in the past. Other authors who have focused on answering the question in some substantial way include Conway (1976), Wild (1994), Billard (1998), and Lindsay et al. (2004). The question is also often asked and answered in elementary textbooks and on the webpages of departments of statistics as well as on those of departments for other disciplines offering training in statistics to their students.

My excuse for using the title again is to explain how I, as editor of the *Annual Review of Statistics and Its Application*, one of the newest journals published by Annual Reviews, view the goals and the coverage of our enterprise. In what follows, I begin by describing what others have said in answer to the question: What is statistics? Then, I briefly review the emergence of statistics as a discipline, from the earliest efforts of John Graunt in the mid-seventeenth century to the present, where I situate statistics in the context of discussions about the future of the field and the role of applications in it.

2. WHAT OTHERS SAY

Most introductory statistical texts begin with an obligatory opening paragraph or possibly a separate box on "What is Statistics?" Here are a few examples:

1. "Statistics is a way of reasoning, along with a collection of tools and methods, designed to help us understand the world" (De Veaux et al. 2006, p. 2).
2. "Statistics is the art of making numerical conjectures about puzzling questions" (Freedman et al. 1978, p. xiii).
3. "Statistics is a set of concepts, rules, and methods for (1) collecting data, (2) analyzing data, and (3) drawing conclusions from data" (Iversen & Gergen 1997, p. 4).
4. "Statistics helps provide a systematic approach for obtaining reasoned answers together with some assessment of their reliability in situations where complete information is unobtainable or not available in a timely manner" (Johnson & Tsui 1998, p. 2).
5. "Statistics is the art and science of gathering, analyzing, and making inferences from data" (Mosteller et al. 1961, p. 2).
6. "Statistics is the art of learning from data. It is concerned with the collection of data, its subsequent description, and its analysis, which often leads to drawing conclusions" (Ross 1996, p. 5).
7. "Statistics is a collection of procedures and principles for gaining and processing information in order to make decisions when faced with uncertainty" (Utts 1996, p. 5).
8. "Statistics is a body of methods for making wise decisions in the face of uncertainty" (Wallis & Roberts 1962, p. 11).

The good news is that these answers overlap considerably and have many words in common. The bad news is that they do not lead us to a simple answer of the form "Statistics is. . . . " An amalgam of the answers above would go something like this: Statistics is both art and science.[1] It is about conjectures, decisions, and inference. It is about data and about learning or reasoning from them. Deming (1943, p. 1) goes further and suggests that "The object of taking data is to provide a basis for action." Statistics is also a set of tools and methods, typically rooted in mathematics

[1]Most statisticians would agree with statistics as science, but many would dispute whether the informal aspects of study design and data analysis referred to by Mosteller et al. (1961) should be described as art. Rather they would claim that these aspects of design and data analysis have now become fully part of the science of statistics.

and probability, that are used to assessess reliability and uncertainty. And, statistics deals with computation because without that, we cannot actually apply the tools and methods. Statistics is a way of thinking about data and working with them to answer substantive questions. As John Tukey once noted, "Doing statistics is like doing crosswords except that one cannot know for sure whether one has found the solution" (Brillinger 2002).

Of course, the definitions of statistics offered above come from introductory textbooks and thus focus on basic methodology and the application of statistics rather than the theory. In a report on the future of statistics drawing on a workshop, Lindsay et al. (2004, p. 38) note that the question has been addressed repeatedly, and they offer the following definition:

> Statistics is the discipline concerned with the study of variability, with the study of uncertainty and with the study of decision-making in the face of uncertainty. . . . It is united through a common body of knowledge and a common heritage. A distinguishing feature of the statistics profession, and the methodology it develops, is the focus on a set of cautious principles for drawing scientific conclusions from data.

We often refer to the consequence of invoking these principles as "statistical thinking," although no clear definition or description of it has been agreed to by most statisticians.

3. FROM OFFICIAL STATISTICS AND POLITICAL ARITHMETIC TO STATISTICS AS A DISCIPLINE

The statistical analysis of data is often traced back to John Graunt and his demographic work on the *Bills of Mortality* in the mid-seventeenth century. The word statistics comes to us via the German word *Statistik*, introduced by Gottfried Achenwall in 1749, which derived from the Latin *status* and is related to the word state (John 1883). He used it to designate the analysis of data about the state, which in English was referred to as political arithmetic. In the eighteenth century, statistics was typically used in the context of data about society, often in what we have come to know as official statistics, but it also had other more specific applications in economics, demography, and politics. These uses of statistics lacked formal methodological techniques for gathering and analyzing data because methods for sample surveys and census taking were in their infancy well into the nineteenth century (Fienberg & Tanur 2001). Statistical practice was also closely linked to physics and astronomy and to the theory of errors. This theory was attributable to many scientists, including Legendre and Gauss, but it also built on earlier work by Laplace and others. For many more details, the reader is referred to Fienberg (1992), Stigler (1986), and Porter (1986), among others.

A key figure in the movement to formalize the gathering and analysis of social statistics was Lambert Adolphe Jacques Quetelet, who trained in mathematics and then in probability and its application. Quetelet's 1835 work, *Sur l'Homme et le Développement de ses Facultés*, *ou Essai de Physique Sociale*, which discussed "the average man," was intended to create a form of social physics, and in it Quetelet analyzed data on a wide variety of social and demographic variables. In his review of Quetelet's book, W.C. Taylor (1836, p. 205) suggested that "The knowledge thus acquired soon assumed a definite form, and statistics, from being a mere assemblage of facts, gradually rose to the dignity of a science, inasmuch as it connected its facts together by a chain of causation." This commentary sounds surprisingly current, particularly in its introduction of the notion causation. Taylor offered several examples of what he meant by causation, for example, in connection with the causes of suicides going from single instances of a relationship to many and the role of interventions in suicide prevention, but his description lacked many of the subtleties of more modern discussions of the topic. Similarly, Taylor meant something very different by the term statistical science, as used in the title of his review essay, than we do today. For Taylor, Quetelet was turning statistics into a science precisely because he trying to infer general laws of human action. In discussing a table compiled by Quetelet on murders in France, Taylor (1836, p. 213) observed:

> It is the business of a statistician to collect and tabulate facts in order to discover the laws of their occurrence; it is no part of his proper duty to investigate their causes. But in the present instance there are other authenticated facts, that seem at least to illustrate a general principle . . . which deserves to be noticed.

Inference takes a very different form in today's statistical science.

The 1830s also saw the creation of statistical associations with the founding of the Royal Statistical Society in 1834 (as the Statistical Society of London) and the American Statistical Association in 1839. Economic statistics dominated both of these societies, and for years, their journals, created in 1838 and 1888, respectively, contained little in the way of formal methodology as we know it today. The International Statistical Institute was created in 1885, following a series of international conferences organized largely through the efforts of William Farr, Florence Nightingale, and Quetelet, to serve as a bridge between countries and to develop standards for government statistics.

The state of government statistics and academic social science in the United States around 1870 is well illustrated through the personage of Francis Amasa Walker, who at age 29 was appointed as the superintendent of the US Census in 1870 (the first post–Civil War census) and who was reappointed in 1880. Walker (1874) helped create the *Statistical Atlas of the United States* to display the census results. His statistical tools were basically maps, graphs, and tabulations. Walker served as the first editor of the *Journal of the American Statistical Association* (1888) and as the president of the American Statistical Association (1883–1897). He taught at Yale University in the 1870s and served as the president of the Massachusetts Institute of Technology from 1881 until his death in 1897. Walker also helped found the American Economic Association in 1885, served as its first president in 1886, and became vice-president of the National Academy of Sciences in 1890 (Anderson 2001).

Near the end of the nineteenth century, the roots of a theory of statistics emerge from the work of Francis Galton and Francis Ysidro Edgeworth and from that of Karl Pearson and George Udny Yule somewhat later. These scientists came to statistics from biology, economics, and social science more broadly, and they developed more formal statistical methods that could be used not just within their fields of interest but across the spectrum of the sciences. Stigler (1986, p. 361) writes about the influence of their efforts as follows:

> The conceptual triumphs of the nineteenth century had been the product of many minds working on many problems in many fields, and one of the most striking of their accomplishments was the creation of a new discipline. Before 1900 we see many scientists of different fields developing and using techniques we now recognize as belonging to modern statistics. After 1900 we begin to see identifiable statisticians developing such techniques into a unified logic of empirical science that goes far beyond its component parts.

Hacking (1990, p. 2) similarly identifies this turning point: "By the end of the century chance had attained the respectability of a Victorian valet, ready to be the logical servant of the natural, biological and social sciences." Pearson also founded the methodological journal *Biometrika*.

4. STATISTICAL METHODS ARE NOW USED ACROSS THE SCIENCES

Building on the pioneering work of Galton, Edgeworth, Pearson, and Yule, other individuals such as Ronald Fisher, Jerzy Neyman, and Egon Pearson helped establish the methodological foundation of the new field by developing experimental designs for the study of agriculture and genetics and more general statistical theory. In a parallel effort, Harold Jeffreys focused on a different philosophical basis and drew on his experiences in geophysics and astronomy. At present, statistics interacts vigorously with astronomy, biology, engineering, geology, medicine and public health, and many social sciences, including political science, law, sociology, psychology, anthropology, archeology, and history. Indeed, we find that statistical methods and ideas are in use in virtually every intellectual field that involves data and quantitative reasoning about them.

A separate movement was afoot in the United States in the 1930s, influenced in part by the efforts of Fisher, Neyman, and Pearson, namely the creation of separate departments of statistics and journals propagating the new mathematical statistics. Agresti & Meng (2013) recently edited a volume that documents the growth of these departments and, in many instances, the tensions between mathematical statistics and work on applied problems. One source of tension was that a generation or more of statisticians grew up in departments of mathematics, resulting in a general view that statistics was part of the mathematical sciences, in some senses joined at the hip. The statistician George Box (2013, p. 18) observed that "[a] serious mistake has been made in classifying statistics as part of the mathematical sciences. Rather it should be regarded as a catalyst to scientific method itself." Box goes on to describe his failed effort to integrate statistics with computer science at the University of Wisconsin, where he founded the department of statistics some 45 years before the emergence of machine learning.

Statistics today consists of a spectrum of activities from foundations, through theory, to methods and data analysis, and some statisticians work in all of these areas. Along this spectrum, statistics draws upon different aspects of mathematics as needed. The discipline also focuses on the philosophy of inductive thinking and on causation. Moreover, statistics involves aspects of computation that have revolutionized both its core methodology and its application. Statistics can be found in departments of statistics, in many other parts of universities, in industry, and in government, often going by different names such as biometrics, chemometrics, econometrics, environmetrics, machine learning, psychometrics, sociometrics, and now data science.

5. THE CURRENT CONTEXT

Let us return to the title of this essay: "What is statistics?" Today, we read and hear much about big data and data science. These terms seem to be constantly in the news, and different groups at universities are rushing to offer courses and even advanced degrees in these topics. But, as others have observed (e.g., Broman 2013), "If you're analyzing data, you're doing statistics. You can call it data science or informatics or analytics or whatever, but it's still statistics."

For me, statistics is what statisticians do and how they think about what they do. That may sound tautological, but there is something essentially different about how statisticians view statistics compared with how most substantive scientists do. This difference does not preclude substantive scientists from being statisticians in both thought and deed. Brown & Kass (2009, p. 107) describe this discrepancy in the following way:

> Different statisticians would use somewhat different words to describe what defines the essential elements of our discipline's approach, but we believe there is general consensus about the substance, which can be stated quite concisely. Statistical thinking uses probabilistic descriptions of variability in (1) inductive reasoning and (2) analysis of procedures for data collection, prediction, and scientific inference.

They go on to note:

> If someone is able to (i) appreciate the role of probabilistic reasoning in describing variation and evaluating alternative procedures and (ii) produce a cutting-edge cross-disciplinary analysis of some data, should we feel comfortable calling that person a statistician? We think so, and we would like to see our profession broaden its perspective to a sufficient degree to make this possible.

Brown & Kass (2009) make a clear distinction between short-term statistical consulting and the in-depth study of an application domain, arguing that a particular domain, neuroscience in their case, may well be the sole focus of a statistician's work. I take a related but somewhat different perspective. For me, the core of statistics arises from the integration of work at the interface of multiple applications, and I find work in different areas both challenging and rewarding. Statisticians develop new methodologies in the context of a specific substantive problem, but they also step back and integrate what they have learned into a more general framework using statistical principles and thinking. Then, they can carry their ideas into new areas and apply variations in innovative ways.

Statistics is not just about the methodology in a particular application domain; it also focuses on how to go from the particular to the general and back to the particular again. R.A. Fisher did this by beginning with the design of experiments in agriculture and the analysis of variance in genetics and then constructing a general theory that he and others brought to bear in many other domains from medicine to education and psychology. A more modern example that I have been associated with has developed a general class of mixed membership models from (*a*) an approach to analyzing text data in documents, known as latent Dirichlet analysis (Blei et al. 2003); (*b*) the grade of membership model for medical diagnosis and the analysis of disability (Erosheva 2003, Erosheva et al. 2007); and (*c*) the STRUCTURE model for the analysis of genetic data (Pritchard et al. 2000). The resulting mixed membership models have subsequently been used for the analysis of network data and for individual-level mixture analysis of other variants (see, e.g., Airoldi et al. 2014). These individual-level mixture models are fundamentally different from traditional cluster models in which a unit can belong to one and only one cluster. In mixed membership models, every unit belongs to the equivalent of every cluster, in essence with probabilities that add to one. The contributions by Blei (2014) and by Erosheva, Matsueda, and Telesca (2014) in this volume describe some of this work.

Among my modern models of statisticians who exemplify this perspective of the interplay between multiple areas of application and of statistical methodology and theory are the late George Box, Frederick Mosteller, and John Tukey, who was often quoted as saying, "The best thing about being a statistician is that you get to play in everyone's backyard" (quoted in Leonhardt 2000). The recent autobiographies by Box (2013) and Mosteller (2010) capture some of the flavor of how the application of statistical techniques is infused with core statistical thinking, as well as that of how the best statistical tools are brought to bear on new substantive problems. That statisticians may need deep disciplinary knowledge in another field does not preclude them from working in multiple areas of application, especially over time.

6. WHAT IS THE *ANNUAL REVIEW OF STATISTICS AND ITS APPLICATION*?

Finally, I return to my purpose in writing this article. In-depth work in areas of application lie at the core of modern statistics, and we hope to promulgate this cross-disciplinary perspective in the *Annual Review of Statistics and Its Application*. Our authors are and will be statisticians in the broad sense described above. They will describe advances in different aspects of statistical methodology, as well as how these advances are being put to use in producing cutting-edge cross-disciplinary analyses.

The landscape of statistical methodology and its application has changed dramatically since Annual Reviews was founded in the 1930s, when statistics as the discipline we know today was in its infancy. Although some topics of interest and core ideas, such as likelihood methods for the analysis of cross-classified categorical data, are the same today as they were then, the tools and perspectives of modern statistics are different from those of the 1930s. In the 1930s, we would have had Bartlett's (1935) analysis of second-order interaction in a $2 \times 2 \times 2$ contingency table applied to an agricultural problem. And today we would expect to see discussions of sparse high-dimensional contingency tables using log-linear models and logistic regression, perhaps in more general form with graphical or generalized linear models and with penalty adjustments for sparseness or with mixed membership models and hierarchical Bayesian structures (see, e.g., Fienberg 2011). Bartlett's solution involved solving a cubic equation, but today our computational tools are extensive and include techniques for identifying degeneracies due to sparseness, Markov chain Monte Carlo algorithms, variational Bayesian approximations, and other forms of scalable algorithms. And the applications might well be to biomedical or public health problems involving genome-wide association studies or to the analysis of social survey data. Many core statistical principles remain the same in the transition from a small contingency table to a massively large one, but the tools and methods continue to grow and adapt. For example, I would argue that despite the technical advances and the computational approaches embedded in more general structures, we still need to think in terms of the basic concepts of odds ratios and conditional independence (partial association) introduced by Yule (1900, 1903), as well as those of ratios of odds ratios as described by Bartlett (1935), when we interpret the results of our analyses of contingency tables. In this journal, we hope to publish reviews that will bring these modern ideas to statisticians and nonstatisticians alike. We also expect to include articles that focus on core statistical principles, lest these be lost in a sea of new methods and computational algorithms.

For Quetelet in 1835, social physics was essentially a relatively crude observational science based on averages, the normal distribution, and graphical displays. Today's social science has become increasingly experimental and methodologically rigorous. We have gone well beyond the caution that "association is not causation," attributed to Pearson (1897), and Yule's (1922) description of "spurious correlation" (cf. Aldrich 1995), focusing instead on structural equation models and statistical methods that attempt to grapple with the delicate adjustments required for causal inference, such as instrumental variables, propensity score matching, and network-based experimental inference. We have gone from using simple graphical representation of trends over time to using group-based trajectories and latent structure to describe behavior. And research on causal inference appears in economics, philosophy, and sociology journals, not only in statistics journals. Again, these topics are ripe for exposition in our review articles.

I have chosen two examples, contingency table analysis and causal inference, but I could have written similarly about the evolution of statistical thinking associated with time series analysis, again going back to early contributions of Yule (see selected papers reproduced in Yule 1971), or about spatial statistics or meta-analysis, beginning with Fisher (1925, 1948). Over time, we hope to provide comprehensive reviews and discussions that cover a broad spectrum of topics within the modern statistical landscape.

A typical modern university has a large number of departments, many of which use and teach statistics, including one or more departments of statistics or biostatistics. This diversity makes keeping track of and coordinating statistical activity across units difficult. University libraries have similar problems with the cataloging of statistics books and methodological journals, which may be physically shelved or stored electronically under mathematics or statistics as well as under biology, business, economics, psychology, sociology, and many other disciplines. Hotelling (1949) wrote about this issue more than 60 years ago, and it continues to affect the disciple of statistics today.

Annual Reviews as an organization reflects this multiplicity of engagement with statistics. In many Annual Reviews journals, one can find articles with major statistical applications and even expository articles about topics in statistical methodology. Some of these latter articles might well have appeared in the *Annual Review of Statistics and Its Application* had it existed at the time they were published. Our articles are intended not only for statisticians but also for subject matter specialists across disciplines who are users of statistical methods and/or methodologists themselves. We want to engage the readers of other Annual Reviews journals and to cross-link relevant review articles in both directions. I hope that at Annual Reviews we will be able to build upon one another's efforts, much as happens in universities with statistics departments of the kind envisioned by Brown & Kass (2009), i.e., focused both on statistical thinking and methodology and on serious cross-disciplinary work.

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

I am especially grateful to Joel Greenhouse, Nancy Reid, and Stephen Stigler, who provided thoughtful comments and criticisms on early drafts of materials for this article.

LITERATURE CITED

Agresti A, Meng X-L, eds. 2013. *Strength in Numbers: The Rising of Academic Statistics Departments in the U.S.* New York: Springer

Airoldi EM, Blei DM, Erosheva EA, Fienberg SE. 2014. *Handbook of Mixed Membership Models.* Boca Raton, FL: Chapman & Hall\CRC Press

Aldrich J. 1995. Correlations genuine and spurious in Pearson and Yule. *Stat. Sci.* 10(4):364–76

Anderson M. 2001. Francis Amasa Walker. In *Statisticians of the Centuries*, ed. CC Hyde, E Seneta, pp. 216–18. New York: Springer-Verlag

Bartholomew DJ. 1995. What is statistics? *J. R. Stat. Soc. A* 158(1):1–20

Billard L. 1998. The role of statistics and the statistician. *Am. Stat.* 524:319–24

Bartlett MS. 1935. Contingency table interactions. *J. R. Stat. Soc. Suppl.* 2:248–52

Blei DM. 2014. Build, compute, critique, repeat: data analysis with latent variable models. *Annu. Rev. Stat. Appl.* 1:203–32

Blei DM, Ng AY, Jordan MI. 2003. Latent Dirichlet allocation. *J. Mach. Learn. Res.* 3(4–5):993–1022

Box GEP. 2013. *An Accidental Statistician: The Life and Memories of George E. P. Box*. New York: Wiley

Brillinger DR. 2002. John W. Tukey: His life and professional contributions. *Ann. Stat.* 30(6):1535–75

Broman K. 2013. Data science is statistics. *The Stupidest Thing: Statistics, Genetics, Programming, Academics*, Apr. 5. **http://kbroman.wordpress.com/2013/04/05/data-science-is-statistics/**

Brown EN, Kass RE. 2009. What is statistics? (with discussion). *Am. Stat.* 63(2):105–23

Conway F. 1976. What is statistics and who should teach it? *J. R. Stat. Soc. A* 139(3):385–88

De Veaux RD, Velleman PF, Bock ED. 2006. *Intro Stats*. Boston: Pearson/Addison Wesley. 2nd ed.

Deming WE. 1943. *Statistical Adjustment of Data*. New York: Wiley

Erosheva EA. 2003. Bayesian estimation of the grade of membership model. In *Bayesian Statistics 7: Proceedings of the Seventh Valencia International Meeting, June 2–6, 2002*, ed. JM Bernardo, MJ Bayarri, JO Berger, AP Dawid, D Heckerman, AFM Smith, M West, pp. 501–10. Oxford: Oxford Univ. Press

Erosheva EA, Fienberg SE, Joutard C. 2007. Describing disability through individual-level mixture models for multivariate binary data. *Ann. Appl. Stat.* 1(2):502–37

Erosheva EA, Matsueda RL, Telesca D. 2014. Breaking bad: two decades of life-course data analysis in criminology, developmental psychology, and beyond. *Annu. Rev. Stat. Appl.* 1:301–32

Fienberg SE. 1992. A brief history of statistics in three and one-half chapters: a review essay. *Stat. Sci.* 7(2):208–25

Fienberg SE. 2011. The analysis of contingency tables: from chi-squared tests and log-linear models to models of mixed membership. *Stat. Biopharm. Res.* 3(2):173–84

Fienberg SE, Tanur J. 2001. Sample surveys, history of. In *International Encyclopedia of the Social & Behavioral Sciences*, ed. NJ Smelser, PB Baltes, pp. 13453–58. Oxford: Elsevier

Fisher RA. 1948. Questions and answers #14. *Am. Stat.* 2(5):30–31

Fisher RA. 1925. *Statistical Methods for Research Workers*. Edinburgh, UK: Oliver & Boyd

Freedman D, Pisani R, Purves R. 1978. *Statistics*. New York: Norton

Hacking I. 1990. *The Taming of Chance*. New York: Cambridge Univ. Press

Hotelling H. 1949. The place of statistics in the university (with discussion). In *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*, pp. 21–49. Berkeley: Univ. Calif. Press

Iversen GR, Gergen M. 1997. *Statistics: The Conceptual Approach*. New York: Springer

John V. 1883. The term "statistics." *J. Stat. Soc. Lond.* 46(4):656–79

Johnson RA, Tsui K-W. 1998. *Statistical Reasoning and Methods*. New York: Wiley

Leonhardt D. 2000. John Tukey, 85, statistician; coined the word 'software.' *New York Times*, July 28. **http://www.nytimes.com/2000/07/28/us/john-tukey-85-statistician-coined-the-word-software.html**

Lindsay BG, Kettenring J, Siegmund DO. 2004. A report on the future of statistics. *Stat. Sci.* 19(3):387–413

Mosteller F. 2010. *The Pleasures of Statistics: The Autobiography of Frederick Mosteller*, ed. SE Fienberg, DC Hoaglin, JM Tanur. New York: Springer

Mosteller F, Rourke REK, Thomas GB Jr. 1961. *Probability with Statistical Applications*. Reading, MA: Addison-Wesley

Pearson K. 1897. On a form of spurious correlation which may arise when indices are used in the measurement of organs. *Proc. R. Soc. Lond. A* 60:489–98

Porter TM. 1986. *The Rise of Statistical Thinking, 1820–1900*. Princeton, NJ: Princeton Univ. Press

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 55(2):945–59

Ross SM. 1996. *Introductory Statistics*. New York: McGraw-Hill

Stigler SM. 1986. *The History of Statistics: The Measurement of Uncertainty Before 1900*. Cambridge, MA: Harvard Univ. Press

Taylor WC. 1836. Objects and advantages of statistical science. *Foreign Q. Rev.* 16:205–29

Utts JM. 1996. *Seeing Through Statistics*. Belmont, CA: Duxbury

Walker FA. 1874. *Statistical Atlas of the United States Based on the Results of the Ninth Census 1870 with Contributions from Many Eminent Men of Science and Several Departments of the Government*. New York: Julius Bien, Lithogr.

Wallis WA, Roberts HV. 1962. *The Nature of Statistics*. New York: Free Press

Wild CJ. 1994. Embracing the "wider view" of statistics. *Am. Stat.* 48(2):163–61

Wilson EB. 1927. What is statistics? *Science* 65(1694):581–87

Yule GU. 1922. *An Introduction to the Theory of Statistics*. London: Griffin. 6th ed.

Yule GU. 1903. Notes on the theory of association of attributes in statistics. *Biometrika* 2(2):121–34

Yule GU. 1900. On the association of attributes in statistics: with illustrations from the material of the childhood society, &c. *Philos. Trans. R. Soc. A* 194:257–319

Yule GU. 1971. *Statistical Papers of George Udny Yule*, selected by A Stuart, MG Kendall. New York: Hafner