# *Preface*

Statistics plays an essential role in diverse fields of human endeavor, including physical science and technology, medicine, public health, the social and behavioral sciences, economics and business. It is the unacknowledged workhorse of scientific analysis and prediction aimed at everything from organ transplants to climate science to financial products. At the same time, the emerging field of "big data" — the vast troves of data generated by advancing technology — suggests future directions of a field that continues to develop.

Yet, the breadth and influence of statistics are largely unknown among the general public. The expository chapters in this volume describe some of the contributions of Canadian statisticians and illustrate the breadth and impact of the field. My hope is that the picture of statistics "in action" they provide will stimulate readers from many backgrounds. Those wishing to find out more about statistics (or statistical science) might consult *Statistics in the 21st Century* (Chapman & Hall/CRC, 2002) and a volume commissioned by the Committee of Presidents of Statistical Societies (COPSS) for its 50th anniversary and the International Year of Statistics, *Past, Present, and Future of Statistical Science* (Chapman & Hall/CRC, 2014).

In the first chapter, Bellhouse and Fienberg review the development of statistics as a discipline in Canada. They note the close relationship with statistics groups in the United States, and the extent to which early generations of Canadian statisticians trained there. The following chapter by Beaumont, Fortier, Gambino, Hidiroglou, and Lavallée describes some of the major contributions to survey methodology made at Statistics Canada, one of the world's premier official statistics agencies.

The next five chapters discuss types of statistical methodology that find application in a wide variety of fields. Ramsay and Hermanussen consider the area known as functional data analysis, and illustrate its use in understanding growth in children. Genest and Nešlehová discuss how to model simultaneous variation in two or more variables and illustrate this methodology in a risk management context involving losses associated with fire insurance claims. Tibshirani describes the "lasso" and related methods for finding important variables among large numbers of factors, with an application to genetics. Rosenthal explains the Metropolis algorithm, a so-called Markov Chain Monte Carlo (MCMC) method originally used in theoretical physics; such algorithms are crucial in statistical applications ranging from artificial intelligence to genetics. Finally, Bingham, Ranjan, and Welch describe computer experiments

in which large scale computer models are used to study complex phenomena such as volcanic eruptions or climatic variation.

The next seven chapters describe applications of statistics in medicine and public health. Two chapters, by Bull, Graham, and Greenwood and by Craiu and Sun, describe how statistical methods are used to look for genetic factors associated with human diseases or traits. Gustafson explains how Bayesian statistical methods can help deal with measurement problems associated with observational health sciences data. Cook shows how statistical models are used to understand data from clinical trials involving cancer patients. Asgharian, Wolfson and Wolfson discuss methods for studying the time that people live after onset of dementia. Schaubel and Kalbfleisch use statistics to address an important issue in kidney transplantation concerning the use of lower versus higher quality organs. Steiner describes statistical methods for monitoring patients following medical treatment, and their use in improving health care.

Chapters 15 and 16 describe two topics of considerable current relevance: the application of statistical methods in finance and in e-commerce, respectively. Rémillard discusses how statistical models are used to price financial products and to assess the associated risks. Zhu explains some of the methods used to decide what products or services are presented to Internet users.

The remaining chapters deal with various scientific areas of vital importance. Chapters by Cowen, Challenger, and Schwarz and by Rivest and Baillargeon describe "mark-recapture" methods for the study of populations that are difficult to access. They illustrate how the methods can be used to estimate the size of populations as diverse as salmon stocks, injection drug users and illegal immigrants. Mills Flemming and Field show how statistical models are used in marine ecology to study the abundance of fish stocks and endangered species. Zwiers, Hegerl, Zhang, and Wen discuss how evidence for climate change is assessed, and how statistics is used to study the impact of human factors. Finally, Kouadio and Newlands describe how satellite data and other new forms of information are combined with statistical models to provide forecasts of crop yields and other agricultural features.

Canadian statisticians have made major contributions to many other areas of statistical theory and methods, as well as to the foundations of statistical inference. In the selection of areas represented here we have tried to emphasize topics of major current importance and to consider what might lie ahead. One strand running through these chapters is the importance of computing to statistical modeling and analysis. A topic of great current interest is so-called "big data," which refers to the volumes of information generated, for example, by the Internet, financial systems and other automated processes. Issues involved with huge amounts of data are discussed in several of the chapters in this book, and we can expect more and more attention to this area in future.

I want to thank Christian Léger, 2012–13 president of the Statistical Society of Canada, for suggesting that a volume of this type be prepared. I also want to acknowledge the help I had reviewing chapters from numerous readers and from the associate editors: David Bellhouse, Hugh Chipman, Richard Cook, Christian Genest, Paul Gustafson, Agnes Herzberg, Nancy Reid, and Doug Wiens. Brian Allen provided good advice and promotion for the volume. Finally, I am most grateful to the managing editor, Christian Genest, who has expertly compiled this volume.

Jerald F. Lawless, PhD, FRSC
Editor