Curriculum Guidelines for Canadian Undergraduate Programs in Statistics

Karen Buro, MacEwan University
Sotirios Damouras, University of Toronto Scarborough
Alison Gibbs, University of Toronto
John Petkau, University of British Columbia
John Sheriff, University of Lethbridge
Jim Stallard, University of Calgary
Bethany White, University of Toronto

Background & Motivation

Statistical science is enjoying a period of rapid growth in both scope and practice. The increasing availability of massive amounts of data and computational power is expanding the purview of Statistics in novel and exciting ways. These developments are also placing new demands on the education and training of future statisticians. For our undergraduate students to be immediately employable as statisticians/data scientists/analysts after graduation, there is a pressing need to update curricula to keep pace with the rapid evolution of the discipline.

Recognizing these challenges, and inspired by a similar endeavour by the American Statistical Association (see http://www.amstat.org/education/curriculumguidelines.cfm), a working group emerged from the 2015 *Advancing Innovation and Scholarship in Statistics Education in Canada* workshop to develop guidelines for undergraduate Statistics curricula in Canada. The guidelines presented here were developed in consultation with various stakeholders (e.g., Statistical Society of Canada membership, alumni and current students, and industry contacts) and are presented in the form of recommended program outcomes and associated suggested topics and skills. They represent a current vision of professional competencies our undergraduate students should possess upon graduation to ensure they are prepared to take on a broad range of professional roles.

Core Areas

To prepare graduates from undergraduate degrees in Statistics for versatile careers, existing and emerging, we have identified four core areas that should be developed through the curriculum:

- *Statistical Methodology*: The statistician's toolkit has also expanded dramatically in recent years. Predictive modelling and observational data have become as prominent in the work that statisticians do as inferential/explanatory methods and experimental data. Graduates from undergraduate degrees in Statistics should be able to select, apply, and assess appropriate statistical methods for their purposes.
- *Computing with Data*: The practical analysis of data is increasingly reliant on large-scale computation. As the functionality and availability of existing technologies improve and new technological tools emerge, Statistics graduates need to be adept at statistical computation and be conscious of new developments and able to adapt accordingly. This demands a solid foundation in computer science training to prepare Statistics graduates to think algorithmically, to program in a high-level language, and to manage and process data proficiently.
- *Statistical Practice*: Graduates from undergraduate degrees in Statistics should be given ample opportunity to apply their knowledge, practice and hone their skills, and develop the attributes of effective practicing statisticians through authentic learning in realistic conditions. They should be able to solve problems and address real-world questions by collecting and analyzing real data, communicating their results effectively, employing best practices and ethical standards in the context of real-world problems.
- *Theoretical Foundations*: Graduates from undergraduate degrees in Statistics should have a mastery of the mathematical tools used in Statistics, and comprehension of the theoretical underpinnings of statistical inference and prediction.

How to Use these Guidelines

The guidelines provide high-level suggestions for undergraduate programs and courses in Statistics (e.g., Statistics Major programs). They are expressed in terms of learning outcomes, and associated topics and skills, grouped based on the four core areas described above. They are not intended to be prescriptive; rather they were written in the hopes that they would be flexible enough to be of value in program reform and course design discussions across institutions and contexts. We see the four core areas as essential, but they need not be equally weighted and they are not presented here in order of importance. Further, these learning outcomes need not be supported exclusively through coursework; for instance, they could be developed and reinforced through a combination of classroom instruction, course projects, consulting/research opportunities, or internships throughout the program. The working group hopes that institutions across Canada will find the guidelines helpful as they plan their specific curricula, based on their academic strengths, faculty complement, resources, and strategic plans.

Undergraduate Program-level Learning Outcomes

1. Statistical Methodology: Upon graduation, students completing undergraduate Statistics programs should be able to appropriately apply, compare, and evaluate methods of collecting and analyzing data. In particular, they should be able to:

- 1.1 Design methods for the collection of appropriate data to answer questions.
- 1.2. Explore data to gain insights into their behavior and characteristics.
- 1.3. Select appropriate models and techniques to address statistical questions.
- 1.4. Implement various statistical methods and draw conclusions from their results.
- 1.5. Assess assumptions and recognize limitations of their data collection and analysis choices.

2. Computing with Data: Upon graduation, students completing undergraduate Statistics programs should be able to implement a substantial statistical analysis project in a programming environment of their choice. In particular, they should be able to:

- 2.1. Store and retrieve data of various types and formats, and manipulate them efficiently.
- 2.2. Explore and effectively visualize and summarize data, and perform statistical modeling, inference and/or prediction.
- 2.3. Write efficient and well-documented code that supports reproducible research, including writing algorithms and conducting simulations.
- 2.4. Scale up computation for handling big data.

3. Statistical Practice: Upon graduation, students completing undergraduate Statistics programs should be able to translate real-world problems into statistical questions, and communicate their findings in an accessible manner. In particular, they should be able to:

- 3.1. Distill statistical questions from the contextual information of the problem at hand.
- 3.2. Recommend appropriate designs of statistical studies to address real-world problems.
- 3.3. Clearly communicate concepts, methods and results to diverse audiences orally, and in writing.
- 3.4. Function effectively in a collaborative and/or interdisciplinary environment.
- 3.5. Demonstrate ethical practice in the conduct of data collection, data handling, statistical analysis and reporting [1].

4. Theoretical Foundations: Upon graduation, students completing undergraduate Statistics programs should possess the necessary theoretical sophistication required to critically appraise and interpret statistical models and methods. In particular, they should be able to:

- 4.1. Apply fundamental results from calculus and linear algebra.
- 4.2. Use probability theory to describe and analyze random phenomena.
- 4.3. Identify important issues in the modeling and analysis of data (e.g. randomness, sources of variation, association vs. causation, model uncertainty) and understand the foundations and implications of different approaches to analyzing data.

[1] In accordance with the Statistical Society of Canada (SSC) Code of Ethical Statistical Practice and the ethics requirements described in the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (TCPS) and the Canadian Council on Animal Care in Science (CCAC) standards.

Suggested Topics and Skills

1. Statistical Methodology: Methods for descriptive, inferential, and predictive data analyses.

- **Data Collection:** Experimental vs observational data, methods for collecting experimental, survey and observational data**.**
- **Descriptive Statistics:** Graphical and numerical summaries of data for exploring different types of variables and their relationships.
- **Modeling:** Modeling relationships of response and explanatory variables (simple, multiple and generalized linear regression), models for multivariate and time-series data, simulation-based methods, nonparametric and machine learning methods for classification and prediction (decision trees, nonparametric regression, neural networks, support vector machines).
- **Inferential Statistics:** Estimation and hypothesis testing for statistical model parameters.

2. Computing with Data: Necessary computational tools for performing data analysis with big and/or complex real-world data.

- **Data Processing:** Tools for accessing and manipulating large structured (SQL) and unstructured (e.g., XML, text, web-page) data.
- **Statistical Software:** Proficiency in at least one language or environment for data analysis (e.g., R, SAS, Python, etc.).
- **Programming:** Basic programming concepts (e.g., control structures, recursion, elementary data structures algorithms, run-time analysis) and computationally intensive methods.
- **Scalability:** Tools for parallel/distributed data storage and processing (e.g., Hadoop, Spark).

3. Statistical Practice: Regular and frequent opportunities for students to practice their statistical knowledge and skills, with a focus on experiential learning.

- **Practical Experience**: Analysis of large, messy, real world data through capstone projects, internships, consulting opportunities, or interdisciplinary research projects.
- **Communication:** Technical writing and oral presentation skills, opportunities for collaboration and team-work with collaborators from both Statistics and other disciplines.
- **Statistical Literacy**: Critical review of statistical reports: assess assumptions, identification strengths/weaknesses, and evaluate the conclusions and impact.
- **Best Practices:** Development and promotion of professional and ethical standards in statistical practice.

4. Theoretical Foundations: Necessary mathematical background for developing and analyzing statistical models and their properties.

- **Calculus and Linear Algebra:** Differentiation and integration in several variables, matrix operations, linear systems, Euclidean spaces and matrix decompositions.
- **Probability:** Discrete and continuous probability distributions in several variables, (in)dependence and conditioning, stochastic processes.
- **Theory of Statistical Inference:** Scientific method, causal vs descriptive inference, predictive modelling, likelihood concepts, frequentist and Bayesian perspectives, and parametric and nonparametric approaches.