American Statistical Association
Draft Guidelines for Undergraduate Programs in Statistical Science

Guidelines Workgroup
(Beth Chance, Steve Cohen, Scott Grimshaw, Johanna Hardin, Tim Hesterberg, Roger Hoerl, Nicholas Horton (chair), Chris Malone, Rebecca Nichols, Deborah Nolan)

May 15, 2014

Draft Statement[1]

Statistics is increasingly important in the modern world, due in part to the proliferation of complex and rich data as well as growing recognition of the role that statistics plays in making evidence-based decisions. At the same time, enrollments in statistics classes have been increasing dramatically. More students are entering college having completed a statistic class, and more students are studying statistics at the college level. There is growing demand for a variety of strong undergraduate programs in statistics to help prepare the next generation of students to make sense of the information around them.

The American Statistical Association (ASA) endorses the value of undergraduate programs in statistical science, both for statistics majors and for students in other majors seeking a minor or concentration. The ASA Board of Directors endorsed a set of guidelines in 2000. But much has changed since then. This document describes updated and expanded guidelines for curricula for undergraduate programs (majors, minors, and concentrations) in statistical science that account for important changes in the field.

Principles

Goals: Undergraduate programs in statistics should equip students with quantitative skills that they can employ and build on in flexible ways. Some students plan graduate work in statistics or other fields, while most others will seek employment after their bachelors degree.

Undergraduate statistics programs should emphasize concepts and tools for working with complex data and provide experience in designing studies and analyzing real data (defined as

[1]this is a preliminary draft: please do not quote or cite. Comments and suggestions welcomed to `nhorton@amherst.edu` by June 15, 2014.

data that has been collected to solve an authentic and relevant problem) that go well beyond the content of a second course in statistical methods. The detailed statistical content of these "data science" skills may vary, but should be integrated with study in computing, data skills, mathematics, and (ideally) a field of application.

Most undergraduate programs are not intended to train professional statisticians, though some graduates may reach this level through work experience and/or further study. Programs should be sufficiently flexible to accommodate varying student goals.

Thinking with Data and Data Science: Statistics students need the capacity to make sense of the staggering amount of information collected in our increasingly data-centered world, and to accurately and effectively communicate these findings. This requires the development of data, computing, and visualization capacities to complement more traditional mathematical and statistical skills.

Previously, undergraduate students with knowledge of statistical software were able to successfully navigate analyses that were given to them. However, with data now taking all shapes and formats, our students require the ability to program in higher level languages as well as fluently interact with databases. The additional need for skills to compute with data in the context of answering a statistical question represents the most salient change since the prior guidelines, and many programs will require considerable creativity to fully integrate these skills into the curriculum. A number of data science topics need to be considered for inclusion into introductory, second, and advanced courses in statistics to ensure that students develop the ability to frame and answer statistical questions with richer supporting data early in their programs, and move towards more dextrous ability to compute with data in later courses.

Flexibility: Institutions vary greatly in the type and breadth of programs they are able to offer, but the ASA believes almost all institutions can provide a level of statistical education that is useful to both students and employers. This document lays out general goals and some specific recommendations. Institutions should adapt these guidelines to meet the needs of their own students, potentially with tracks within a single program. Each institution will also need to regularly review their programs over time, to reflect new developments in this fast-moving field.

Teaching statistical theory: Understanding the theoretical underpinnings of statistical methods is a vital component of modern statistical practice. While we do not presume to specify how the ideal statistical theory (mathematical statistics) course should be structured, we do believe that aspects of the traditional probability/inference sequence, with its emphasis on large sample size approximations and lists of distributions, does not reflect current statistical practice. A modern statistical theory course might, for example, include more work on computer intensive methods and non-parametric modeling. This course should provide students with an overview of statistics and statistical thinking that builds on what is provided in their introductory statistics courses. It may be useful to incorporate computing, data-related, and communication components in this class. If included early on in a student's program, it will help to provide a solid foundation for future courses and experiential opportunities.

Relationship with other disciplines: Though statistics requires mathematics for the development of its underlying theory, statistics is distinct from mathematics and uses many non-mathematical skills. Few statistics students need the mathematics to derive classical statistical formulas, many of which are often superseded by computational approaches that are more accurate and better facilitate understanding. (Students planning graduate study in statistics, however, need more background in mathematics and theoretical statistics, in addition to strong computing skills.)

More important than the mathematics skills for most students is the ability to work with data, to learn to ask questions and understand the right questions to ask, to use a variety of computational approaches to manipulate data and find answers, and to communicate results in a comprehensible and correct fashion. Therefore, it is essential that faculty developing statistics curricula and teaching courses be trained in statistics and experienced in working with data (mathematical expertise is not a substitute).

Working with data requires extensive computing skills far beyond what was needed in the past. In addition to facility with professional statistical analysis software, students require the ability to access and manipulate data in various ways, to program, and to utilize algorithmic problem solving. Relationships with computer science and allied disciplines are increasingly important.

Relationship with masters programs in statistics: There are a number of differences between the learning outcomes of masters programs and those at the bachelors level, primarily related to level. There is a presumption that masters graduates are statisticians and as such are able to move directly towards accreditation or other professional recognition. Graduates from undergraduate programs in statistics are more generally employed as analysts or similar positions which utilize a number of statistical skills. Both levels of graduates are needed to help address the shortage of workers with the skills to make evidence-based decisions informed by data.

The ASA workgroup report on guidelines for masters programs made a number of recommendations which we reproduce here:

1. Graduates should have a solid foundation in statistical theory and methods,
2. Programming skills are critical and should be infused throughout the graduate student experience,
3. Communication skills are critical and should be developed and practiced throughout graduate programs,
4. Collaboration, teamwork, and leadership development should be part of graduate education,
5. Students should encounter non-routine, real problems throughout their graduate education, and
6. Internships, co-ops or other significant immersive work experiences should be integrated into graduate education.

We note that many of these recommendations also apply to undergraduate programs. There may be opportunities for programs with both graduate and undergraduate programs in statistics to facilitate access by undergraduate students to curricular innovations which address these recommendations.

Relationship with high school and community college courses in statistics: The dramatic growth of the number of students completing the Advanced Placement course in Statistics combined with the augmented role for statistics as part of the Common Core State Standards have increased the exposure of the discipline at the high school level. Institutions need to re-evaluate their introductory courses to account for these changes.

At the same time, the number of students studying introductory statistics courses at two-year (community) colleges has increased to more than 134,000 per year (larger than the total enrollment in calculus classes at this level, up from a previous ratio of ten calculus sections per statistics section, back in the 1960's). This shift reflects that statistics is a universal discipline, not just needed for a handful of students, but required of a majority of disciplines.

Anecdotal evidence suggests that a number of statistics majors are transferring in from community colleges. A key question is how to facilitate this transfer and ensure that students can successfully undertake preliminary coursework and general education requirements in preparation to complete a statistics degree at a four-year institution. Further efforts are needed to streamline articulation agreements with community colleges as well as support faculty development and curricular development at two-year colleges.

Learning outcomes and assessment: There is a growing awareness of the importance of learning outcomes (a detailed list of what a student is expected to know, understand, and demonstrate after completing a program) and assessment of these learning outcomes. Many internal and external groups (such as accreditors, legislators, parents, and students) are calling upon institutions to demonstrate accountability by defining learning goals and objectives at the program level (in addition to the course level) and devising strategies for assessing whether these goals and objectives are being met.

Assessments can be structured in a number of ways. They can be direct (e.g., tests, projects) or indirect (e.g., surveys, focus groups). For higher order thinking skills, which encompass much of a statistics program, assessments should be authentic, open-ended, and complex. A good assessment plan will include indication of where (e.g., which courses, experiences) students are expected to develop the skills, and when they are expected to be introduced to, practice, and master the skills. Further work is needed to identify appropriate learning outcomes and assessment strategies for statistics programs.

Other resources: These guidelines are intended to provide an overview of a principled approach to ensuring that statistics students have the appropriate skills and ability to solve complex data-based problems in the world. Some key resource materials and an annotated bibliography are available at XX WEBSITE TO BE ESTABLISHED XX.

Skills Needed

Effective statisticians at any level display an integrated combination of skills that are built upon a mathematical foundation but which encompass statistical theory, statistical application, computation, data manipulation, and communication. Beginning students cannot be assumed to be able to fully comprehend these myriad connections, and an appropriate developmental progression is required to develop mastery. These skills need to be introduced, developed, and assimilated throughout a student's academic program, beginning with introductory courses, and reinforced in later classes, and ideally, a capstone experience.

We have not specified a mininum number of classes (or equivalent) expected in each area, though we detail some examples of courses and programs which provide preparatory, introductory, intermediate, and advanced skill development with an integrated approach. Ideally there should be no information silos.

Programs should provide sufficient background in the following areas:

Statistical: Graduates should have training and experience in statistical reasoning, in designing studies (including practical aspects), in exploratory analysis of data by graphical and other means, and in a variety of formal inference procedures.

Programming: Graduates should have knowledge and capability in a programming language, along with the ability to think algorithmically, to tackle programming/scripting tasks, and design and carry out simulation studies.

Data-related skills: Working with data requires more than basic computing skills. Programs should require prowess with a professional statistical software package. Graduates should have demonstrated skill in data management and manipulation, knowledge of database technologies, as well as project management and reproducible analysis tools.

Mathematical foundations: Undergraduate major programs should include study of probability and statistical theory, along with the prerequisite mathematics, especially calculus and linear algebra. Programs for non-majors may require less study of mathematics. Programs preparing for graduate work in theoretical statistics may require additional mathematics (e.g., mathematical analysis and advanced calculus).

Communication: Graduates should be expected to write clearly, speak fluently, and have developed skills in collaboration and teamwork as well as the ability to organize and manage projects. These skills need to dovetail with their technical and statistical knowledge. Undergraduate majors in statistics will often be hired into positions as analysts, where they need to be able to understand and communicate findings regarding uncertainty and risk. Many academic programs may not fully prepare students for these important workplace skills.

Substantive area: Because statistics is a methodological discipline, statistics programs should include some depth in an area of application.

Curriculum Topics for Undergraduate Degrees

The approach to teaching should:

1. Emphasize authentic real-world data and substantive applications,
2. Present data in a context that is both meaningful to students and indicative of the science behind the data,
3. Include experience with statistical computing early and often,
4. Encourage synthesis of theory, methods, and applications,
5. Offer frequent opportunities to refine communication skills, tied directly to instruction in technical skills,
6. Be models of the correct application of statistics, and
7. Highlight linkages between topics, concepts, and capacities being taught.

Statistical Topics

Statistical thinking begins with a problem and explores data to answer key questions. Students need a deep understanding of fundamental topics as well as exposure to a variety of topics and methods, though it is not feasible or appropriate to attempt a comprehensive overview of the entire field at the undergraduate level.

1. Statistical theory (e.g., distributions of random variables, point and interval estimation, hypothesis testing, decision theory, Bayesian methods)
2. Graphical data analysis methods (e.g., advanced visualization techniques, smoothing/kernel estimation, exploratory data analysis tools as well as tools to create visual displays for presentations)

3. Statistical modeling (e.g., variety of regression models, categorical data, non-parametric methods, survival analysis, diagnostics, incomplete data, causal inference)
4. Design of studies (e.g., random assignment, replication, blocking, confounding ; analysis of variance, fixed and random effects, diagnostics in experiments, complex surveys)
5. Data mining, multivariate methods, clustering, and predictive analytics
6. Simulation and resampling based approaches (e.g., bootstrapping, permutation tests)

Programming Topics

1. Programming concepts (e.g., ability to break down problems in a step-by-step solution, algorithms, data representations, structured programming, and basic software engineering)
2. Programming for simulation studies, cross-validation, and other tools commonly employed in the modern use of statistics
3. Modern methods for large scale problems (e.g., text mining, parallel computing, and precursors to "big data")

Data-Related Topics

Graduates need the ability to flexibly and correctly manage and restructure data to answer a variety of questions. Such skills underpin any analysis.

1. Facility with professional statistical software appropriate for a variety of tasks
2. Understanding of data manipulation and management techniques
3. Experience with reproducible analysis and project management tools and workflows designed to minimize errors
4. Knowledge of database systems and their interface to statistical tools
5. Knowledge of strategies for assessing and ensuring data quality as part of data preparation

Mathematical foundation topics

1. Calculus (integration and differentiation) through multivariable calculus

2. Applied linear algebra (e.g., matrix manipulations, linear transformations, projections in Euclidean space, eigenvalues/eigenvectors and important matrix decompositions)

3. Conditional probability, random variables, multivariate and joint continuous as well as discrete distributions

4. Use of simulation (empirical) exploration to complement analytic results

5. Emphasis on connections between concepts in probability and linear algebra and their applications in statistics

Communication Topics

While communication skills are important in any field, they are particularly necessary for statistics, and complement technical knowledge to ensure that results are made available in an accessible and correct manner.

1. Effective technical writing and presentations

2. Teamwork and collaboration

3. Planning for data collection

4. Data management (real-life, messy, data integrity)

5. Ability to use technical skills to undertake analyses and communicate results and conclusions concisely and effectively

6. Professional conduct and ethics

Electives

There are many electives that might be included in a statistics major. As resources will vary among institutions, the identification of what will be offered is left to the discretion of individual units.

Practice

Undergraduate curricula must provide ample opportunities to practice the work of being a statistician. Whenever possible, the undergraduate experience should include one or more opportunities for internships, senior-level "capstone" courses, consulting experiences, or a combination of these. These and other ways to practice statistics in context should be included in a variety of venues in an undergraduate program. The completion of such practice requirements in statistics can help ensure that graduates have the skills to work as a practicing statistician.

Curriculum Topics for Minors or Concentrations

It is challenging to develop the capacity to be able to analyze data in the manner that we describe within the constraints of an undergraduate program that might include 10–12 courses. The issues are even striking for minor programs or concentrations, which typically feature a much smaller number of courses as part of their requirements. In some cases, however, statistics minors or concentrations for quantitatively oriented students in fields such as biology, business, and behavioral and social science or those planning to teach at the K–12 level may be more feasible than a full statistics major, and institutions need to design their programs to ensure a core set of skills. These programs will necessarily be more varied than major programs.

The core of a minor or concentration in statistics should consist of the following:

1. General statistical methodology (statistical thinking, descriptive, estimation, testing, etc.)

2. Statistical modeling (simple and multiple regression, diagnostics, etc.)

3. Experience designing a simple survey or experiment

4. Facility with professional statistical software, along with data management skills (e.g. data cleaning and consistency checking, exposure to reproducible research tools) tools

5. Multiple experiences analyzing data and communicating results

The number of credit hours for minors or concentrations will depend on the institution. Additional topics to consider include applied regression, design of experiments, statistical computing, data science, theoretical mathematical statistics, categorical data analysis, time series, Bayesian methods, probability, database management, high performance computing, and a capstone experience. Courses from other departments with significant statistical content might be allowed to count toward a statistics minor or concentration, though the content of such courses must differ substantially from the others. A capstone or other integrative experience is particularly useful for minors or concentrations.