

**American Statistical Association  
Undergraduate Guidelines Workgroup**

# Curriculum Guidelines for **Undergraduate Programs** in Statistical Science



*Promoting the Practice and Profession of Statistics®*

# ACKNOWLEDGMENTS

The American Statistical Association undergraduate guidelines working group was convened by ASA President Nathaniel Schenker in the spring of 2013. Members included Beth Chance, Steve Cohen, Scott Grimshaw, Johanna Hardin, Tim Hesterberg, Roger Hoerl, Nicholas Horton (chair), Chris Malone, Rebecca Nichols, and Deborah Nolan. We greatly appreciate the many members of the community who provided feedback on earlier drafts of these guidelines.

# CONTENTS

Executive Summary .....	4
Introduction.....	5
Background and Guiding Principles .....	6
Skills Needed.....	9
Curriculum for Statistics Majors .....	11
Curriculum Topics for Minors or Concentrations.....	14
Additional Points .....	15

These guidelines were endorsed by the American Statistical Association Board of Directors on November 15, 2014. A copy of the guidelines and related resources can be found at [www.amstat.org/education/curriculumguidelines.cfm](http://www.amstat.org/education/curriculumguidelines.cfm).

# EXECUTIVE SUMMARY

The American Statistical Association endorses the value of undergraduate programs in statistics as a reflection of the increasing importance of the discipline. We expect statistics programs to provide sufficient background in the following core skill areas: statistical methods and theory, data management, computation, mathematical foundations, and statistical practice. Statistics programs should be flexible enough to prepare bachelor's graduates to either be functioning statisticians or go on to graduate school.

The widely cited McKinsey report states that “by 2018, the United States alone could face a shortage of 140,000 to 190,000 people with deep analytical skills as well as 1.5 million managers and analysts with the know-how to use the analysis of Big Data to make effective decisions.” A large number of those will be at the bachelor's level. The number of bachelor's graduates in statistics has increased by more than 140% since 2003 (21% from 2012 to 2013).

Much has changed since the previous guidelines were disseminated in 2000. The 2014 guidelines reflect changes in curriculum and suggested pedagogy. Institutions need to ensure students entering the work force or heading to graduate school have the appropriate capacity to “think with data” and to pose and answer statistical questions.

## Key points

**Increased importance of data science.** Working with data requires extensive computing skills. To be prepared for statistics and data science careers, students need facility with professional statistical analysis software, the ability to access and wrangle data in various ways, and the ability to perform algorithmic problem solving. In addition to more traditional mathematical and statistical

skills, students should be fluent in higher-level programming languages and facile with database systems.

**Real applications.** Data should be a major component of statistics courses. Programs should emphasize concepts and approaches for working with complex data and provide experiences in designing studies and analyzing non-textbook data.

**More diverse models and approaches.** Students require exposure to and practice with a variety of predictive and explanatory models in addition to methods for model-building and assessment. They must be able to understand issues of design, confounding, and bias. They need to know how to apply their knowledge of theoretical foundations to the sound analysis of data.

**Ability to communicate.** Students need to be able to communicate complex statistical methods in basic terms to managers and other audiences and to visualize results in an accessible manner. They must have a clear understanding of ethical standards. Programs should provide multiple opportunities to practice and refine these statistical practice skills.

These guidelines are intended to be flexible while ensuring that programs provide students with the appropriate background and necessary critical thinking and problem solving skills to thrive in our increasingly data-centric world. Programs are encouraged to be creative with their curriculum to provide a synthesis of theory, methods, computation, and applications.

A copy of the guidelines and related resources can be found at <https://goo.gl/Ncjj3v> (last updated November 15, 2014).

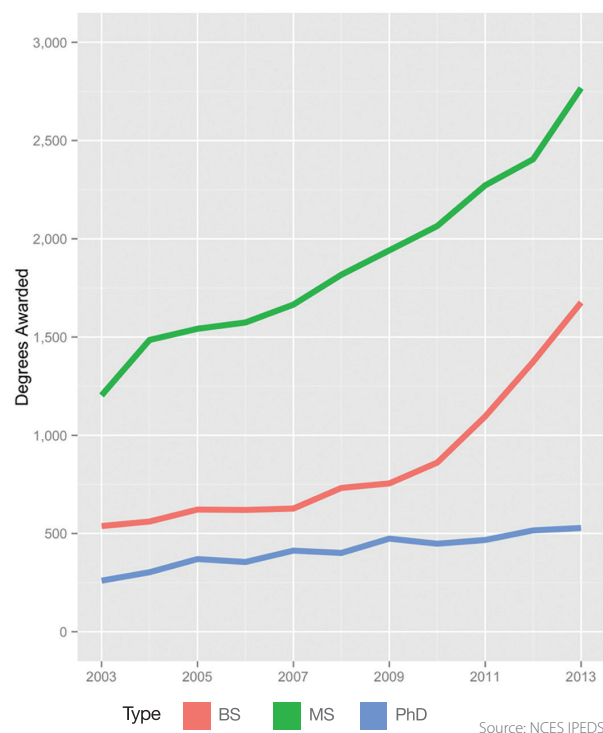
# INTRODUCTION

Statistics is an increasingly important discipline, spurred by the proliferation of complex and rich data and the growing recognition of the role statistical analysis plays in making evidence-based decisions. Enrollments in statistics classes have been increasing dramatically. More students are entering college having completed a statistics class, and more students are studying statistics at the college level. Although the number of bachelor's graduates in statistics is still relatively small in absolute terms (1,656 according to IPEDS<sup>1</sup>), this number has increased markedly from 2003, when only 673 statistics undergraduate degrees were conferred<sup>2</sup>.

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 widely cited McKinsey & Company report stated that “by 2018, the United States alone could face a shortage of 140,000 to 190,000 people with deep analytical skills as well as 1.5 million managers and analysts with the know-how to use the analysis of Big Data to make effective decisions.”<sup>3</sup> While some of these new workers will need graduate training, much of the demand is expected to be at the bachelor's level<sup>4</sup>.

The American Statistical Association (ASA) endorses the value of undergraduate programs in statistical science, both for statistics majors and students in other majors seeking a minor or concentration<sup>5</sup>. Much has changed since the previous ASA guidelines, which were approved in 2000<sup>6</sup>. This document describes updated and expanded guidelines for curricula for undergraduate programs (majors, minors, and concentrations) in statistical science that account

TRENDS IN STATISTICS DEGREES AWARDED



for important changes in the field. We lay out general goals and specific recommendations identified during our deliberations from 2013–2014.

We begin by discussing principles that informed our thinking, then consider skills students should develop in their courses, and finally summarize key curriculum topics.

# NOTES

**1** Data are from IPEDS (Integrated Post-Secondary Education Data System Completions Survey, [ncesdata.nsf.gov/webcaspar](http://ncesdata.nsf.gov/webcaspar)) through 2013, where first and second majors were counted in biostatistics, statistics, mathematical statistics and probability, statistics [other], and mathematics and statistics [other].

**2** See, for example, <http://magazine.amstat.org/blog/2014/09/01/degrees> and <http://magazine.amstat.org/blog/2013/05/01/stats-degrees>. The University of California/Berkeley (n=143) was the largest producer in 2013, with Purdue University a close second (n=135). Other institutions with 40 or more graduates in 2013 included University of Illinois/Urbana-Champaign, University of California/Davis, University of Minnesota, University of Michigan, and University of California/Los Angeles. There may be substantial undercounting, since students completing a mathematics major with a

concentration in statistics are not included in these numbers.

**3** See [www.tinyurl.com/mckinsey-nextfrontier](http://www.tinyurl.com/mckinsey-nextfrontier).

**4** See [www.maa.org/programs/faculty-and-departments/ingenious](http://www.maa.org/programs/faculty-and-departments/ingenious).

**5** We focus primarily on majors, since the development of a deep understanding of statistical science and associated computational and data-related skills requires extensive study. We also describe key points related to minor programs and similar types of concentrations or tracks through other majors.

**6** See [www.amstat.org/education/curriculumguidelines.cfm](http://www.amstat.org/education/curriculumguidelines.cfm) for the 2000 guidelines and related resources.

7 The K–12 GAISE guidelines ([www.amstat.org/education/gaise](http://www.amstat.org/education/gaise)) define statistical problem solving as an investigative process that involves four components: (1) Formulate questions (clarify the problem at hand, then formulate one (or more) questions that can be answered with data); (2) Collect data (design a plan to collect appropriate data, then employ the plan to collect the data); (3) Analyze data (select appropriate graphical and numerical methods, then use these methods to analyze the data); and (4) Interpret results (interpret the analysis, then relate the interpretation to the original question). It should be emphasized that this process is rarely sequential. In addition, see Wild and Pfannkuch (1999) “Statistical thinking in empirical enquiry,” *International Statistical Review*, 67(3):223–248 and Pfannkuch and Wild (2000) “Statistical thinking and statistical practice: Themes gleaned from professional statisticians,” *Statistical Science*, 15(2):132–152.

8 There is a need for additional continuing professional development for instructors and revisions to the graduate curricula that will prepare future instructors. Because many faculty teaching statistics do not have a graduate degree in statistics, there is a need for creative approaches to ensure they have appropriate background (see, for example, the 2014 ASA/MAA guidelines for teaching statistics, <http://magazine.amstat.org/blog/2014/04/01/asamaaguidelines>).

9 See [www.tinyurl.com/cupm2004](http://www.tinyurl.com/cupm2004) and the 2015 guidelines at [www.maa.org](http://www.maa.org).

# BACKGROUND AND GUIDING PRINCIPLES

**The scientific method and its relation to the statistical problem solving cycle:** Undergraduates need practice using all steps of the scientific method to tackle real research questions. All too often, undergraduate statistics majors are handed a “canned” data set and told to analyze it using the methods currently being studied. This approach may leave them unable to solve more complex problems out of context, especially those involving large, unstructured data. The statistical analysis process involves formulating good questions, considering whether available data are appropriate for addressing the problem, choosing from a set of different tools, undertaking the analyses in a reproducible manner, assessing the analytic methods, drawing appropriate conclusions, and communicating results<sup>7</sup>. Students need practice developing a unified approach to statistical analysis and integrating multiple methods in an iterative manner. Instructors need appropriate background in applied statistics and the statistical problem solving cycle to be able to effectively teach these courses<sup>8</sup>.

This scientific approach to statistical problem solving is important for all data analysts, not just undergraduate statistics majors or minors. It needs to start in the first course and be a consistent theme in all subsequent courses. Often, there is more than one appropriate way to address a research question. Students need to see that the discipline of statistics is more than a collection of unrelated tools (or methods); it is a general approach to problem solving using data. Undergraduates need to develop judgment to assess approaches and verify assumptions,

including nonstatistical justifications (subject matter knowledge) for evaluating a research conclusion. Students need to be aware of possible limitations, to assess when a more complex analysis is warranted, and to decide when to reformulate the question.

**Real applications:** The Committee on the Undergraduate Program in Mathematics Curriculum Guide from 2004 reinforced the importance of real applications and data analysis<sup>9</sup>. They stated:

[T]he analysis of data provides an opportunity for students to gain experience with the interplay between abstraction and context that is critical for the mathematical sciences major to master. Experience with data analysis is particularly important for majors entering the workforce directly after graduation, for students with interests in allied disciplines, and for students preparing to teach secondary mathematics.

We concur and recommend that a focus on data be a major component of introductory and advanced statistics courses and that students work with authentic data throughout the curriculum. Institutions should ensure that modern applied statistics courses are available early in the curriculum. These courses are particularly relevant for strong mathematics students and have the potential to recruit students into statistics and other mathematical sciences programs. It is also essential that faculty developing statistics curricula and teaching courses be trained in statistics and experienced



in working with data<sup>10</sup>. Instructors need additional materials that feature real applications, including curated data sets, sample syllabi, and other resources.

More generally, undergraduate statistics programs should emphasize concepts and approaches for working with complex data and provide experiences in designing studies and analyzing real data (defined as data that have been collected to solve an authentic and relevant problem) that go well beyond the content of a second course in statistical methods<sup>11</sup>. The detailed statistical components of these problem solving skills may vary, but should be tightly integrated with study in statistics, data wrangling, computing, mathematics, and, ideally, a field of application<sup>12</sup>.

**Focus on problem solving:** Undergraduate programs in statistics should equip students with problem solving skills they can effectively apply, build on, and extend over time. They should teach principles that will allow graduates to ask questions, assess their work, and learn new ideas as needed. Many bachelor's graduates seek employment immediately after their degree<sup>13</sup>. Some flexibility in the undergraduate curriculum is needed, as the appropriate skill-set for those seeking employment immediately upon graduation may differ from those seeking admission into doctoral programs in statistics.

**The increasing importance of data science:** Statistics students need to make sense of the staggering amount

of information collected in our increasingly data-centered world and to manage data, analyze it accurately, and communicate findings effectively<sup>14</sup>. This capacity has been elegantly described by Diane Lambert of Google as the ability to “think with data.”<sup>15</sup> Although a formal definition of data science is elusive, we concur with the StatsNSF committee statement that data science comprises the “science of planning for, acquisition, management, analysis of, and inference from data.”<sup>16</sup>

With increasingly large data, the relative importance of statistical topics changes. Methods that find patterns and relationships in high-dimensional data become more important, as do methods to avoid bias from available data. Model assessment remains critical, while statistical significance is less central.

In previous decades, it was often sufficient for undergraduate majors in statistics who had knowledge of statistical software to successfully navigate analytic tasks assigned to them. Students now need facility with professional statistical analysis software<sup>17</sup>, the ability to access and manipulate data in various ways, and the ability to use algorithmic problem solving. They need to learn to pose relevant questions (to gain insight), use a variety of computational approaches to extract meaning from data, judge data quality, assess their models and methods, and communicate results in a comprehensible and correct fashion. With data now taking all shapes and formats, statistics majors need to be able to program in higher-level languages<sup>18</sup> and fluently interact with database systems. The additional need to think with data—in the context of answering a statistical question—represents the most salient change since the prior guidelines were endorsed in 2000<sup>19</sup>.

Adding these data science topics to the curriculum necessitates developing data, computing, and visualization capacities that complement more traditional mathematically oriented statistical skills<sup>20</sup>.

**10** The 2014 ASA/MAA guidelines for teaching introductory statistics (<http://magazine.amstat.org/blog/2014/04/01/asamaaguidelines>) make the same recommendation. Mathematical expertise is not a substitute.

**11** There is not a single definition of what is appropriate as a second course in statistics, and a number of options can be found at many institutions. No matter how innovative the approach, we believe it is not possible to develop a comprehensive understanding of the range of key statistical concepts after only two courses.

**12** 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 institutions.

**13** Data from a survey of graduates from California Polytechnic State University, San Luis Obispo (Melissa Bowler, unpublished senior project) found that 60% of bachelor's graduates eventually completed a graduate degree, but often not until many years in the work force. A study of graduates from Eastern Kentucky University found similar results (see Kay and Costello, *JSM* 2014, and Costello and Kay (2002), “Where do all of the undergraduate statistics majors go?,” *STATS*, 34:10–13). Additional studies of graduates and their early career profiles would be valuable for the community.

**14** See, for example, the National Academies report “The Mathematical Sciences in 2025,” [www.nap.edu/catalog.php?record\\_id=15269](http://www.nap.edu/catalog.php?record_id=15269).

**15** See also the ASA report “Discovery with Data: Leveraging Statistics with Computer Science to Transform Science and Society,” [www.amstat.org/policy/pdfs/BigDataStatisticsJune2014.pdf](http://www.amstat.org/policy/pdfs/BigDataStatisticsJune2014.pdf), and “Thinking with Data: How to Turn Information into Insights” by Shron (2007).

**16** See [www.nsf.gov/attachments/130849/public/Stodden-StatsNSF.pdf](http://www.nsf.gov/attachments/130849/public/Stodden-StatsNSF.pdf).

**17** There are numerous examples of software packages that can be used for introductory statistics courses, including JMP, Minitab, R/ RStudio, SAS, SPSS, and Stata.

**18** We define this as a programming environment that supports abstraction from a specification to the computer (e.g., hides many aspects of the underlying computational environment), such as Python, R, or SAS.

**19** See Nolan and Temple Lang (2010) “Computing in the statistics curriculum,” *The American Statistician*, 64(2):97–107 for an overview of key curriculum topics in data science.

**20** See also the white paper by Hardin et al., “Data Science in the Statistics Curricula: Preparing Students to ‘Think with Data.’”

**21** It will be increasingly important to engage with faculty involved in computing education (e.g., members of the Association for Computing Machinery Special Interest Group in Computer Science Education, SIGCSE) to learn of their experiences and approaches to teaching “computational thinking.”

**22** This is by no means new advice. The 35-year-old ASA report “The Training of Statisticians for Industry” (1980, *The American Statistician*, 34(2):65–75) describes skills for an effective industrial statistician, particularly the role of communication taught in conjunction with technical topics. Then and now, if students develop a clear understanding of basic statistical theory that allows them to select, use, and assess a model, it is more likely they can learn to effectively use other approaches that they were not exposed to in college (or did not exist before they graduated).

**23** This is also true for more theoretical master’s programs in statistics and applied mathematics.

**24** See the white paper “Roadmap for Smaller Schools” by Hoerl.

---

## The main goal of our recommendations is to ensure undergraduate statistics students remain useful in a world with increasingly more complex data.

---

**Creative approaches to new curricular needs:** Many programs will require considerable creativity to fully integrate additional data-related and statistical practice skills into the curriculum. Relationships with allied disciplines that teach applied statistics, and with computer science, will become increasingly important. 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 rich supporting data early in their programs, and move towards dexterous ability to compute with data in later courses<sup>21</sup>. To make room, some traditional topics will need to be dropped from the core curriculum. We do not attempt to specify which topics are central, and which could be covered in electives or dropped entirely. Given that most undergraduate statistics majors enter the workforce as analysts where data-skills are primary, we suggest that helping them to master a smaller set of methods, rather than a comprehensive laundry list, is likely to be more useful to them in the long-term<sup>22</sup>.

The main goal of our recommendations is to ensure undergraduate statistics students remain useful in a world with increasingly more complex data. If we don’t prepare them to learn new techniques and work with various forms of data, it will be difficult for them to compete for jobs. We need to pay attention to the core foundations of statistical thinking and practice without shying away from increasingly important data science skills.

We recognize the hurdles and challenge at all ends of the spectrum to ensure that students are provided with modern statistical experiences. For programs that are already implementing computational courses, faculty should be encouraged to share resources, make course content available, and help train the next generation of teachers and scholars. For programs that are unable to implement an entire major program, we suggest that missing topics or skills be added to classes in the current curriculum. Additional co- and extra-curricular experiences that enhance the formal statistics curriculum should be embraced and encouraged.

**Relationship with mathematics:** Though the practice of statistics requires mathematics for the development of its underlying theory, statistics is distinct from mathematics and requires many nonmathematical skills. Few undergraduate statistics students need the mathematics used to derive classical statistical formulas, many of which are often superseded by computational approaches that are more accurate and may better facilitate understanding. Theoretical/mathematical and computational/simulation approaches are complementary, each helping to clarify understanding gained from the other. Students planning doctoral study in statistics need a strong background in mathematics and theoretical statistics in addition to strong computing skills<sup>23</sup>.

**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. Programs should be sufficiently flexible to accommodate varying student goals. Institutions should adapt these guidelines to meet the needs of their students, potentially with tracks within a single program<sup>24</sup>. Each institution should regularly review their programs to reflect new developments in this fast-moving field.



# SKILLS NEEDED

Effective statisticians at any level need to master an integrated combination of skills built upon statistical theory, statistical application, data management, computation, mathematics, and communication. It cannot be assumed that beginning students fully comprehend these myriad connections, and an appropriate developmental progression is required to obtain mastery. Providing students with a strong foundation in statistical methods and theory is critically important for all undergraduate programs in statistics. These skills need to be introduced, supported, and reinforced throughout a student's academic program, beginning with introductory courses and augmented in later classes<sup>25</sup>. Such scaffolded exposure helps students connect statistical concepts and theory to practice.

We have not specified a minimum number of classes (or equivalent) expected in each area, though programs need to provide preparatory, introductory, intermediate, and advanced skill development with an integrated approach. Ideally, there should be many opportunities for topics and concepts that cut across numerous classes to be referenced and integrated in multiple places within the curriculum. Statistics programs should provide majors with sufficient background in the following areas<sup>26</sup>:

**Statistical methods and theory:** Graduates should be able to design studies, use graphical and other means to explore data, build and assess statistical models, employ a variety of formal inference procedures (including resampling methods), and draw appropriate scope of conclusions from the analysis<sup>27</sup>. They need knowledge and experience applying a variety of statistical methods, assessing their appropriateness, and



communicating results. They need a foundation in theoretical statistics principles for sound analyses.

**Data management and computation:** Graduates should be facile with professional statistical software and other appropriate tools for data exploration, cleaning, validation, analysis, and communication. They should be able to program in a higher-level language<sup>28</sup>, to think algorithmically, to use simulation-based statistical techniques, and to undertake simulation studies<sup>29</sup>. Graduates should be able to manage and marshal data, including

**25** Ideally such a program would culminate with capstone and/or internship experiences.

**26** We anticipate departments can use these high-level categories to define program outcomes.

**27** Our enumeration of key statistical skills is intentionally short, since these are likely most familiar to the statistics community. More detail is provided regarding computation and data-related skills because they have not played as large part in the undergraduate statistics curriculum in the past. We reiterate, however, that statistical fundamentals are at the core, with the data-related skills supporting the ability to analyze and interpret complex data.

**28** This capacity includes the ability to write functions and use control flow in a variety of languages and tools such as Python, R, SAS, or Stata. Facility with spreadsheet tools such as Excel is useful for a variety of other purposes, but is not ideal as a programming or reproducible analysis environment.

**29** The capacity to undertake and interpret simulation studies as a way to complement analytic understanding and/or check results will be increasingly useful in the workplace.

**30** Many graduate programs strongly recommend at least a year of mathematical analysis and/or advanced calculus, while other upper-level mathematics courses such as “Stochastic Processes,” “Graph Theory,” “Differential Equations,” “Optimization,” “Combinatorics,” and “Algebraic Statistics” also may be helpful.

**31** One possible model to develop project management skills can be found in the University of California/Berkeley CS169 “Software Engineering” course, which incorporates a substantive project with external customers structured with four two-week-long iterations with those clients ([www.aramandofox.com/2012/05/10/about-uc-berkeley-cs169-software-engineering](http://www.aramandofox.com/2012/05/10/about-uc-berkeley-cs169-software-engineering)). Another approach would be to incorporate project management into capstone experiences.

**32** There is pedagogical value in having students practice communication to identify gaps in their understanding. In addition, communication skills need to dovetail with students’ technical and statistical knowledge: Excellent communication of inappropriate or incorrect analyses is counterproductive.

**33** Data from a survey of graduates from California Polytechnic State University, San Luis Obispo (Melissa Bowler, unpublished senior project) was used to generate a listing of current jobs for  $n=62$  graduates from Cal Poly San Luis Obispo’s undergraduate statistics program. She found that 12 had “statistic” in the title (e.g., Statistician, Senior Statistical Analyst, Statistical Programmer I) while 20 had “analy” in the title (e.g., Data Analyst, Marketing Data Analyst, Research Analyst, Business Systems Analyst). We suspect many of those with “Statistician” in their job title completed a higher degree. Better data on the outcomes of graduates would benefit the profession as a whole.

**34** The interaction of statisticians with subject-matter professionals is a key characteristic of the discipline, as statistics is increasingly a “team sport.” This is particularly important at the planning stage of a study or project. Graduates need to translate subject-matter objectives to statistical plans and analyses that mesh with and are capable of meeting those objectives. Depth in a substantive area provides the capability to engage in this manner.

**35** See, for example, the analytics co-major in the department of statistics at Miami University, [www.miamioh.edu/cas/academics/departments/statistics/academics/majors/analytics-comajor](http://www.miamioh.edu/cas/academics/departments/statistics/academics/majors/analytics-comajor).



joining data from different sources and formats and restructuring data into a form suitable for analysis. Their statistical analyses should be undertaken in a well-documented and reproducible way.

**Mathematical foundations:** Graduates should be able to apply mathematical ideas from linear algebra and calculus to statistics, and to set up and apply probability models. Minor programs will generally require less study of mathematics. Students preparing for doctoral work in statistics should usually complete additional mathematics courses<sup>30</sup>.

**Statistical practice:** Graduates should be expected to write clearly, speak fluently, and construct effective visual displays and compelling written summaries. They should demonstrate ability to collaborate in teams and to organize and manage projects<sup>31</sup>. They should be able to

communicate complex statistical methods in basic terms to managers and other audiences and visualize results in an accessible manner<sup>32</sup>. Undergraduate majors in statistics often will be hired into analyst positions, where they need to be able to understand and communicate statistical findings<sup>33</sup>.

**Discipline-specific knowledge:** Students should be able to apply statistical reasoning to domain-specific questions. This capacity includes translating research questions into statistical questions and communicating results appropriate to different disciplinary audiences. Because statistics is a methodological discipline, statistics programs should encourage study in a substantive area of application<sup>34</sup>. Some programs might include a required second major, co-major<sup>35</sup>, minor, or sequence of related courses to accompany the completion of a statistics degree.

# CURRICULUM FOR STATISTICS MAJORS

## Statistical Methods and Theory

Statistical thinking begins with a problem and explores data to answer key questions. Undergraduate statistics students need a deep understanding of fundamental concepts as well as exposure to a variety of topics and methods<sup>36</sup>, including the following:

- Statistical theory (e.g., distributions of random variables, likelihood theory, point and interval estimation, hypothesis testing, decision theory, Bayesian methods, and resampling<sup>37</sup>)
- Exploratory data analysis approaches and graphical data analysis methods<sup>38</sup>
- Design of studies (e.g., random assignment, random selection, data collection, and efficiency<sup>39</sup>) and issues of bias, causality, confounding<sup>40</sup>, and coincidence
- Statistical models (e.g., variety of linear and nonlinear parametric, semiparametric, and nonparametric regression models<sup>41</sup>; model building and assessment; multivariate methods; and statistical and machine learning techniques<sup>42</sup>)

## Data Wrangling and Computation

Undergraduate statistics majors need facility with computation to be able to handle increasingly complex data and sophisticated approaches to analyze it. Graduates need the ability to manage and restructure



**36** Given how quickly the discipline of statistics is changing, it is not feasible or appropriate to attempt a comprehensive overview of the entire field at the undergraduate level, and we do not attempt it.

**37** Resampling methods (including bootstrapping and permutation tests) are widely applicable to many problems at multiple levels of the curriculum. See the white paper “What Teachers Should Know About the Bootstrap: Resampling In the Undergraduate Statistics Curriculum” by Hesterberg.

**38** These topics include advanced visualization techniques, smoothing/kernel estimation, spatial methods (see manuscript by Christou “Enhancing the Teaching of Statistics Using Spatial Data”), and mapping. We also note the value of visualization early in the analysis process to identify errors and anomalies.

**39** Other important topics include blocking, stratification, survey sampling, and adaptive designs.

**40** Issues of confounding and causal inference are central to the discipline of statistics. There are many settings in which a randomized experiment cannot be undertaken. To avoid pitfalls of drawing conclusions from observational data, students need a clear understanding of principles of statistical design and tools to assess and account for the possible impact of other measured and unmeasured variables.

**41** These topics will generally include many of the following: simple and multiple linear regression, generalized linear models, generalized additive models, time series, mixed models, survival analysis, spatial analysis, regression trees, model selection, diagnostics, cross-validation, and regularization.

**42** We note the growing use of machine learning methods (which also arise in analytics) to make predictions about future events. See Breiman (2001), “Statistical modeling: The two cultures,” *Statistical Science*, 16(3):199–231; Harville (2014), “The need for more emphasis on prediction: A ‘non-denominational’ model-based approach,” *The American Statistician*, 68(2):71–92 and related discussion; and the white papers on a model data science course (Baumer) and “Data Science in the Statistics Curricula: Preparing Students to ‘Think with Data’” (Hardin et al.).

**43** See Zhu et al. (2013) “Data acquisition and pre-processing in studies on humans: What is not taught in statistics classes,” *The American Statistician*, 67(4):235–241, which includes a series of skills: (1) get to know the study; (2) assess the validity of variable coding; (3) assess data entry accuracy; (4) perform data cleaning; and (5) edit identified data errors.

**44** Although we acknowledge that Microsoft Excel is a common platform for data exchange, we do not recommend it as a primary analysis environment.

**45** Appropriate environments could include R, Python, and SAS, complemented by tools including shell scripts and knitr.

**46** Futschek (2006) defines algorithmic thinking as a set of abilities related to constructing and understanding algorithms: (1) the ability to analyze a given problem; (2) the ability to precisely specify a problem; (3) the ability to find the basic actions that are adequate to the given problem; (4) the ability to construct a correct algorithm to a given problem using basic actions; (5) the ability to think about all possible special and normal cases of a problem; and (6) the ability to improve the efficiency of an algorithm. Futschek, G. (2006). “Algorithmic thinking: The key for understanding computer science,” in R. Mittermeir (Ed.), *Informatics Education—The Bridge Between Using and Understanding Computers* (Vol. 4226, pp. 159–168). Berlin/Heidelberg: Springer. We consider this to be a necessary, but not sufficient component of “computational thinking.”

**47** We define structured programming as the ability to use functions and control structures (e.g., “for” loops).

**48** This recommendation is consistent with the efforts of Conrad Wolfram and the Computer-Based Math initiative, [www.computerbasedmath.org](http://www.computerbasedmath.org) and [www.tinyurl.com/ted-wolfram](http://www.tinyurl.com/ted-wolfram). The incorporation of these tools may be particularly valuable at the bachelor’s level, since students will generally have less technical knowledge (and need to be able to simulate to generate insights and/or check analytic results).

**49** Students should develop the capacity to manipulate formats such as CSV, JSON (JavaScript Object Notation, a data-interchange format that is easy to read, parse, and generate; see Nolan and Temple Lang (2014) *XML and Web Technologies for Data Sciences with R*), XML, databases (see, for example, Ripley (2001) “Using databases with R,” *R News*, 1(1):18–20 and Wickham (2011) “ASA 2009 Data Expo,” *Journal of Computational and Graphical Statistics*, 20(2):281–283), and text data. Because many faculty were not trained in these technologies, continuing education in this area needs to be made a priority.

**50** We are not prescriptive regarding which technologies are incorporated into the curriculum, as long as they are sufficiently flexible and powerful. Many undergraduate statistics students develop expertise in environments such as R/RStudio, Python, and SAS.

**51** Multivariate calculus is recommended.

**52** Markov chains are a useful topic for undergraduate majors in statistics.

**53** This linkage includes topics such as the delta method. In addition, many students might benefit from exposure to modeling and simulation in their mathematics courses as a way to reinforce their computational skills.

data. Such skills underpin strategies for assessing and ensuring data quality as part of data preparation and are a necessary precursor to many analyses<sup>43</sup>.

- Use of one or more professional statistical software environments<sup>44</sup>
- Data management using software in a well-documented and reproducible way<sup>45</sup>, data processing in different formats, and methods for addressing missing data
- Basic programming concepts (e.g., breaking a problem into modular pieces, algorithmic thinking<sup>46</sup>, structured programming<sup>47</sup>, debugging, and efficiency)
- Computationally intensive statistical methods (e.g., iterative methods, optimization, resampling, and simulation/Monte Carlo methods)<sup>48</sup>
- Use of multiple data tools<sup>49</sup>, so graduates are not wedded to one and are better able to learn new technologies<sup>50</sup>

## Mathematical Foundations

The study of mathematics lays the foundation for statistical theory. Undergraduate statistics majors should have a firm understanding of why and when statistical methods work. They should be able to communicate in the language of mathematics and explain the interplay between mathematical derivations and statistical applications.

- Calculus (e.g., integration and differentiation)<sup>51</sup>
- Linear algebra (e.g., matrix manipulations, linear transformations, projections in Euclidean space, eigenvalues/eigenvectors, and matrix decompositions)



- Probability (e.g., properties of univariate and multivariate random variables, discrete and continuous distributions)<sup>52</sup>
- Emphasis on connections between concepts in these mathematical foundations courses and their applications in statistics<sup>53</sup>

## Statistical Practice

Strong communication skills complement technical knowledge and are particularly necessary for statisticians; graduates need technical skills to perform analyses and communication skills to understand clients’ needs and then effectively discuss results and conclusions. Important practical skills include the following:



- Effective technical writing, presentation skills, and visualizations
- Teamwork and collaboration
- Ability to interact with and communicate with a variety of clients and collaborators

Undergraduate curricula must provide ample opportunities to practice the work of being a statistician. The completion of such requirements in statistics can help ensure that graduates have the necessary skills to work as practicing statisticians.

Ethical issues should be incorporated throughout a program<sup>54</sup>. Whenever possible,

the undergraduate experience should include opportunities for internships<sup>55</sup>, senior-level capstone courses<sup>56</sup>, consulting experiences, research experiences, or a combination<sup>57</sup>. These and other ways to practice statistics in context should be included in a variety of venues in an undergraduate program.

### Pedagogical Considerations

The approach to teaching this curriculum should model the correct application of statistics<sup>58</sup>:

- Emphasize authentic real-world data and substantive applications related to the statistical analysis cycle<sup>59</sup>
- Develop flexible problem solving skills
- Present problems with a substantive context that is both meaningful to students and true to the motivating research question
- Include experience with statistical computing and data-related skills early and often<sup>60</sup>
- Encourage synthesis of theory, methods, computation, and applications
- Provide opportunities to work in teams
- Integrate training in professional conduct<sup>61</sup> and ethics
- Offer frequent opportunities to refine communication skills, tied directly to instruction in technical statistical skills
- Incorporate regular assessment to provide authentic feedback.

**54** See the whitepaper “Ethics and the Undergraduate Curriculum,” by Cohen, and “Seeing Through Statistics” (Chapter 26, *Ethics in Statistical Studies*), Utts (2015), which includes topics such as the ethical treatment of human and animal participants, assurance of data quality, appropriate statistical analyses, and unbiased reporting of results.

**55** See the white paper “Undergraduate Internships in Statistics” by Cohen.

**56** See the white paper “Capstones in the Undergraduate Statistics Curricula” by Malone.

**57** A number of innovative programs have been created in recent years to address the need to provide undergraduate statistics students with authentic experiences posing and answering statistical questions. These include DataFest ([http://chance.amstat.org/2013/09/classroom\\_26-3](http://chance.amstat.org/2013/09/classroom_26-3)), Explorations in Statistics Research (see the draft manuscript by Nolan et al.), the Summer Institutes in Biostatistics ([www.nhlbi.nih.gov/funding/training/redbook/sibsweb.htm](http://www.nhlbi.nih.gov/funding/training/redbook/sibsweb.htm)), and other Research Experiences for Undergraduates (REUs).

**58** Just watching instructors analyze data is insufficient. Students need repeated experiences undertaking analysis of real-world data. It is also important that instructors have a history of such experiences (see the 2014 ASA/MAA guidelines for teaching statistics, <http://magazine.amstat.org/blog/2014/04/01/asamaaguidelines>).

**59** While the GAISE college report ([www.amstat.org/education/gaise](http://www.amstat.org/education/gaise)) focuses on the introductory statistics course, many of its tenets are broadly applicable for the principled teaching of statistics.

**60** Our experience has been that programs that require work with real data in the first year or two tend to be able to offer more substantive real experiences (e.g., advanced data analysis or capstones) in later years.

**61** The American Statistical Association issued a statement on continuing professional development ([www.amstat.org/education/cpd.cfm](http://www.amstat.org/education/cpd.cfm)). Statisticians are encouraged to undertake continuing professional development: (1) in methodology and practice, by keeping abreast of new techniques and theory, staying connected with best practice, growing in areas not previously studied (or refreshing forgotten material), and gathering ideas and direction for future research; (2) in technology, by learning about new computational techniques and software tools and by staying on top of trends in technology and new sources of data that are creating major new opportunities for statisticians; (3) in subject matter needed for successful collaboration with other disciplines, to strengthen the interdisciplinary contributions and capabilities of statisticians; and (4) in career success factors such as communication, leadership, and influence skills, which are vital to the impact of individual contributions and the visibility of our profession.

**62** See Cannon et al. (2001), “Guidelines for undergraduate minors and concentrations in statistical science,” *Journal of Statistics Education*, 10(2), [www.amstat.org/publications/jse/v10n2/cannon.html](http://www.amstat.org/publications/jse/v10n2/cannon.html).

**63** There is a pressing need for additional K–12 teachers with the capacity to teach the Common Core State Standards for Mathematics. See the forthcoming ASA SET (Statistical Education of Teachers) report for specific guidance.

**64** A minor in mathematical statistics also could be considered, but it may be challenging to ensure students develop sufficiently strong computational and data-related skills. A concern is that an emphasis on probability and inference may leave these students less prepared for the job skills expected by employers.

**65** A capstone also might include experience in another content area (e.g., health, education, business, sociology, or biology). See also the white paper “Capstones in the Undergraduate Statistics Curricula” by Malone.

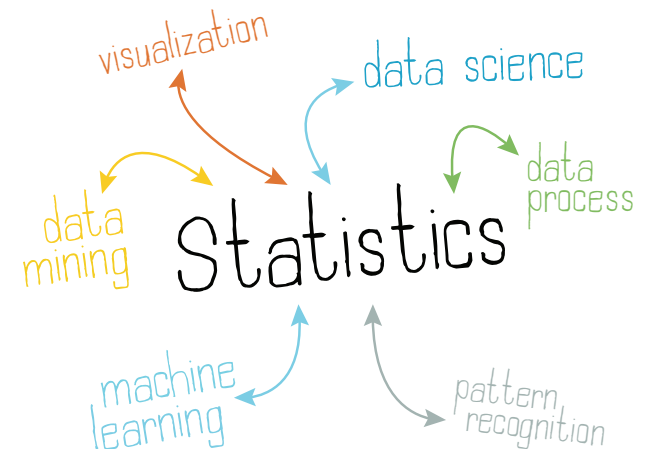
# CURRICULUM TOPICS FOR MINORS OR CONCENTRATIONS

It is challenging to develop the capacity to be able to analyze data in the manner we describe within the constraints of an undergraduate program that might include 10–12 courses. These issues are even more difficult to address for minor programs or concentrations, which typically feature a much smaller number of courses as part of their requirements<sup>62</sup>.

In some cases, however, statistics minors or concentrations for quantitatively oriented students in fields such as biology, mathematics, 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<sup>63</sup>. Institutions need to design such programs to ensure graduates possess a core set of useful 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:

- General statistical methodology (e.g., statistical thinking, descriptive statistics, graphical display, estimation, testing, resampling)
- Statistical modeling (e.g., simple and multiple regression, confounding, diagnostics)
- Facility with professional statistical software, along with data management skills
- Multiple experiences analyzing data and communicating results

The recommendations for minors and concentrations focuses on statistical fundamentals, data technologies, and communication and is intended to ensure students



develop significant data-related skills, understanding of key statistical concepts, and perspective on the field of statistics<sup>64</sup>. The number of credit hours for minors or concentrations will depend upon the institution.

Additional topics to consider include applied regression, design of experiments; statistical computing; data science; theoretical statistics; categorical data analysis; time series; Bayesian methods; probability; database systems; and a capstone, internship, or similar integrative experience<sup>65</sup>. Ethics is another key topic to integrate into these courses. For many students, a methods course in an application area might be an appropriate option. Courses from other departments with substantial statistical content might be allowed to count toward a statistics minor or concentration.

# ADDITIONAL POINTS

**Relationship with high-school and community college courses in statistics:** The dramatic growth of the number of students completing the Advanced Placement Statistics course<sup>66</sup> and the augmented role for statistics as part of the Common Core State Standards for Mathematics have increased the exposure of the discipline at the high-school level<sup>67</sup>. As a result, colleges and universities may need to re-evaluate their introductory courses<sup>68</sup>.

The number of students studying introductory statistics courses at two-year (community) colleges has increased to more than 137,000 per year<sup>69</sup> (larger than the total enrollment in calculus classes at this level, up from a previous ratio of 10 calculus sections per statistics section in the 1960s). This shift reflects the belief that statistics is a universal discipline, not just needed for a handful of students, but required for a number of disciplines and recommended for many others.

Anecdotal evidence suggests many statistics majors are transferring to universities from community colleges<sup>70</sup>. A key question is how to facilitate this transfer and ensure students can successfully undertake preliminary coursework and general education requirements prior to completing a statistics degree at another institution. Further efforts are needed to streamline articulation agreements with community colleges and to support faculty development and curricular development at two-year colleges<sup>71</sup>.

**Relationship with master's programs in statistics:** Graduates from undergraduate programs in statistics are generally employable as analysts or in similar positions that use a number of statistical skills. In addition, a bachelor's degree can and should be considered an



attractive option as a liberal arts degree. Both bachelor's and master's graduates are needed to help address the shortage of workers with the skills to make evidence-based decisions informed by data.

There are differences between the learning outcomes of master's programs and bachelor's programs related to level, breadth, and depth<sup>72</sup>. There has been the presumption that master's graduates are statisticians and

**66** In 1997 (the first year it was offered), a total of 7,667 students took the Advanced Placement Statistics exam. This number increased to 98,033 in 2007 and has increased to more than 185,000 in 2014, making it one of the top 10 largest AP exams.

**67** See [www.corestandards.org/Math/Content/HSS/introduction](http://www.corestandards.org/Math/Content/HSS/introduction) for an overview of statistics and probability topics at the high-school level.

**68** A number of innovative approaches have been suggested for this problem. One approach is to consider a year-long introductory statistics course at the undergraduate level, in which students who have completed Advanced Placement Statistics would begin in the second semester. Offering a sequence of courses (e.g., Applied Statistics I and Applied Statistics II) would facilitate integration of additional topics that are not feasible within the syllabus of a single course.

**69** See the Conference Board of the Mathematical Sciences 2010 report ([www.ams.org/profession/data/cbms-survey/cbms](http://www.ams.org/profession/data/cbms-survey/cbms)).

**70** Data from University of California/Berkeley indicate these numbers are large and increasing (Deb Nolan, personal communication).

**71** See the white paper "The Key Role of Community Colleges to Support the Undergraduate Teaching of Statistics" by Horton et al. and the California online student-transfer information website, [www.assist.org/web-assist/welcome.html](http://www.assist.org/web-assist/welcome.html). Revision of lower-division introductory statistics courses to introduce computing and data science skills will facilitate articulation in the future. A major challenge will be faculty development for instructors in two-year colleges.

**72** The ASA Guidelines for Master's Programs in Statistics (<http://magazine.amstat.org/blog/2013/06/01/preparing-masters>) recommend: (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.

**73** This may be due to undergraduate programs in statistics being rare until recently. The growth in availability of data science jobs, many of which just require a bachelor's, may change this impression.

**74** Most undergraduate programs are not intended to train accredited (professional) statisticians, though some graduates may reach this level through work experience or further study. The American Statistical Association allows graduate statisticians not yet eligible for accreditation because of a lack of experience to be designated GStat ([stattrak.amstat.org/2014/05/01/gstat](http://stattrak.amstat.org/2014/05/01/gstat)). We recommend the development of a similar pathway for those with training at the undergraduate level.

**75** In particular, it may be feasible for master's students with a BS in statistics to get their MS in statistics in a substantially shorter time, or undertake more advanced course work as part of their master's degree.

**76** See [www.amstat.org/sections/educ/MathStatObsolete.pdf](http://www.amstat.org/sections/educ/MathStatObsolete.pdf) for details.

**77** See the white paper by Chance and Peck "From Curriculum Guidelines to Learning Objectives: A Survey of Five Statistics Programs."

**78** They should be "authentic." See, for example, Gould (2010) "Statistics and the modern student," *International Statistical Review*, 78(2):297–315 and Brown and Kass (2010) "What is statistics?," *The American Statistician*, 63(2):105–110 and associated discussion and rejoinder.

undergraduates are not<sup>73</sup>. We disagree. Bachelor's programs should prepare students to be practicing statisticians<sup>74</sup>. Programs with both master's and bachelor's programs can provide bachelor's students access to master's courses. Five-year master's programs (in which students simultaneously receive a bachelor's and master's degree) may be an attractive option for many students.

The growing number of statistics graduates at the bachelor's level may have implications for the structure of and content for master's programs<sup>75</sup>. Further efforts to assist with professional development and continuing education are needed to help ensure that bachelor's graduates have the requisite skills to stay engaged with new developments in the field of statistics.

**Teaching the theoretical underpinnings of statistics:** 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 course (often called "Mathematical Statistics" or "Statistical Inference") 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 fully capture current statistical practice. A lively panel discussion from JSM 2003 raised many relevant issues<sup>76</sup>. A modern statistical theory course might, for example, include work on computer-intensive methods and non-parametric modeling. Such a course should provide students with an overview of statistics and statistical thinking that builds on 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. Because the traditional mathematical statistics course requires probability, which

in turn often necessitates three semesters of calculus as a prerequisite, students often take the theoretical statistics course late in their programs. This sequence precludes other upper-level applied statistics courses building on this important theoretical foundation. Some institutions have been successful in providing students with earlier access to the theoretical underpinnings of the discipline. For example, Brigham Young University splits the traditional probability course into two courses: one with a focus on probability and discrete random variables (taught at the sophomore level) and a more advanced course (with a focus on continuous random variables and additional advanced topics in inference).

**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<sup>77</sup>. 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 and projects) or indirect (e.g., surveys and focus groups). For higher-order thinking skills, which encompass much of a statistics program, assessments should be relevant, open-ended, and complex<sup>78</sup>. A sound assessment plan will include indication of where (which courses?, which 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.



## NEXT STEPS

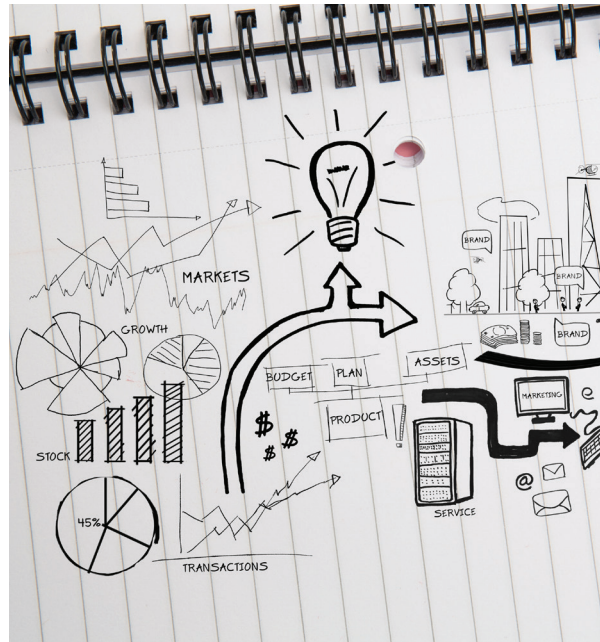
Through our process and discussions, a number of issues arose that we believe merit further exploration over the coming years, but were outside our purview, including the following:

**Faculty development:** The American Statistical Association strategic plan describes the importance of faculty development. Efforts to create and share additional activities, projects, sample syllabi, and model courses will be useful for faculty teaching this curriculum.

**Engagement with two-year colleges:** Community colleges are a large, growing, and increasingly important component of the United States higher education system. Additional efforts are needed to coordinate statistics instruction at the two-year college level, raise the profile of statistics majors at these institutions, and facilitate articulation agreements for transfer to four-year institutions.

**Surveys of graduates and employers:** Better information on the career paths of the growing number of undergraduate statisticians in the work force and surveys of employers would help to guide future curricular changes.

**Certification/accreditation pathway:** The ASA now provides an entry level pathway for master's level statisticians without sufficient work experience to prepare for professional accreditation. While the ASA professional statistician accreditation requires an advanced degree, we believe that there may be other types of certification or accreditation that might be appropriate for workers with a bachelor's degree.



**Multiple pathways for introductory statistics:** This might be an opportunity for the ASA to lead an effort to reassess curricula for a variety of introductory statistics courses. This effort might include delineating model courses for students at two-year colleges, students who have completed AP Statistics, and/or those planning to major in statistics.

**Periodic review:** While we have endeavored to provide a flexible yet specific document, the fast-changing nature of the discipline of statistics will necessitate a review of these curriculum guidelines. Undertaking such a regular review at least every five-eight years would be warranted.

We encourage the wider statistical and mathematical sciences community to explore next steps for each of these items.

## CLOSING

These guidelines are intended to provide an overview of a principled approach to ensure that undergraduate statistics majors have the appropriate skills and ability to tackle complex and important data-focused problems. Additional resource materials and an annotated bibliography are available at [www.amstat.org/education/curriculumguidelines.cfm](http://www.amstat.org/education/curriculumguidelines.cfm).





