Using the Guidelines to Develop a New Undergraduate Program

Stacey Hancock
University of California, Irvine
stacey.hancock@uci.edu
Overview

• Current status of undergraduate programs in Statistics
• Using the guidelines to build an undergraduate degree program
• Development of Data Science undergraduate major at UCI
• Data Science vs. Statistics undergraduate programs
• Challenges and future considerations
Undergraduate Statistics Degrees

“Statistics – the science of learning data – is the fastest-growing STEM undergraduate degree in the United States over the last four years.” Amstat News, 1 March 2015

<table>
<thead>
<tr>
<th>Field</th>
<th>% Increase from 2010</th>
<th># of 2013 Degrees</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistics</td>
<td>95.1</td>
<td>1,656</td>
</tr>
<tr>
<td>Computer Info Tech Admin. &amp; Mgmt.</td>
<td>92.0</td>
<td>1,985</td>
</tr>
<tr>
<td>Environmental/Environmental Health Engineering</td>
<td>86.4</td>
<td>1,195</td>
</tr>
<tr>
<td>Mathematics &amp; Statistics, Other</td>
<td>80.0</td>
<td>216</td>
</tr>
<tr>
<td>Computer Programming</td>
<td>77.2</td>
<td>466</td>
</tr>
</tbody>
</table>

Source: Data from the National Center for Education Statistics; Analysis by the American Statistical Association

130 colleges and universities offer a bachelor’s degree in Statistics.
<table>
<thead>
<tr>
<th>Institution</th>
<th>2003</th>
<th>2013</th>
</tr>
</thead>
<tbody>
<tr>
<td>University of California – Berkeley</td>
<td>30</td>
<td>143</td>
</tr>
<tr>
<td>Purdue University</td>
<td>9</td>
<td>135</td>
</tr>
<tr>
<td>University of Illinois at Urbana-Champaign</td>
<td>8</td>
<td>67</td>
</tr>
<tr>
<td>University of Michigan – Ann Arbor</td>
<td>20</td>
<td>55</td>
</tr>
<tr>
<td>University of California – Davis</td>
<td>7</td>
<td>53</td>
</tr>
<tr>
<td>University of Minnesota – Twin Cities</td>
<td>13</td>
<td>50</td>
</tr>
<tr>
<td>University of California – Los Angeles</td>
<td>0</td>
<td>50</td>
</tr>
<tr>
<td>Iowa State University</td>
<td>6</td>
<td>37</td>
</tr>
<tr>
<td>Brigham Young University – Provo</td>
<td>10</td>
<td>35</td>
</tr>
<tr>
<td>Pennsylvania State University</td>
<td>5</td>
<td>33</td>
</tr>
</tbody>
</table>
ASA Guidelines: Key Points

• Increased importance of data science – the ability to “think with data”
• Real applications as a major component available early in the curriculum
• More diverse models and approaches
• Ability to communicate
ASA Guidelines: Skills

- Statistical methods and theory
- Data manipulation and computation
- Mathematical foundations
- Statistical practice
- Discipline-specific knowledge
Topics vs. Skills

What *topics* should an undergraduate program in statistics cover?

**vs.**

What should a student who completes a major in statistics be able to *do*?
Learning Objectives

• “…statements of what a student is expected to know, understand and/or be able to demonstrate after completion of a process of learning.” (Kennedy, 2007)
• Avoid words such as “know” or “understand”, and use more active words such as “design”, “apply” or “assess”.
• Aim for 5-10 outcomes at the program level.
## Curriculum Mapping

### Learning Objectives

<table>
<thead>
<tr>
<th>Year 4</th>
<th>LO1 Understand foundational mathematical principles</th>
<th>LO2 Understand foundational computer science principles</th>
<th>LO3 Understand foundational statistical principles</th>
<th>LO4 Understand the principles of statistical computing</th>
<th>LO5 Develop solutions for real-world data analysis problems</th>
<th>LO6 Demonstrate effective communication skills</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICS 51</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stats 115</td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Stats 170A</td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Stats 170B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
</tbody>
</table>

### Required Courses
Curriculum Mapping

- Replace the “x”s in each cell with I, D, or M for:
  - **Introduce**: Learn about key ideas and concepts at a general or basic level.
  - **Develop**: Gain additional information related to the outcome; start to synthesize key ideas and skills.
  - **Master**: Demonstrate ability to perform the outcome with a reasonable level of independence and sophistication.

- Each LO should be introduced, developed and mastered at least once across multiple courses.
- Each course should support at least one (ideally more than one) LO.
UCI Department of Statistics

- Formed in 2002
- Ten full-time faculty members as of fall 2015
- MS and PhD graduate degrees in Statistics
- Undergraduate Statistics minor
- Undergraduate Data Science major (starting fall 2015!)
- Undergraduate Statistics major included in our future plans
UCI Department of Statistics

- Donald Bren School of Information and Computer Sciences
- All three departments (Statistics, Computer Science, Informatics) housed in Donald Bren Hall
- Close working relationships between the departments
B.S. in Data Science at UCI

- UCI Data Science Initiative (datascience.uci.edu)
- Two Statistics (Stacey Hancock and Jessica Utts) and two Computer Science (Padhraic Smyth and Mike Carey) faculty developed the major proposal
B.S. in Data Science at UCI

- Dual emphasis in statistics and computer science
- Core statistics courses include:
  - Introductory statistics, exploratory data analysis and statistical computing, one-year applied statistics sequence, one-year probability and mathematical statistics sequence, Bayesian statistics
- Core computer science courses include:
  - Introductory and intermediate programming, data structures, logic and discrete mathematics, algorithms, software engineering, machine learning and data mining, data management, information visualization
- First year seminar in data science
- Two quarter senior capstone project course
Educational Goals

• Provide students with a foundation in mathematical and statistical aspects of data analysis;

• Provide students with a foundation in the general principles of computer science;

• Teach students how to utilize their knowledge of statistical and computing principles to develop algorithms, and software for solving real-world data analysis problems;

• Provide students with practical experience in applying their knowledge of theories, methods, and tools, to a variety of data analysis problems;

• Teach students how to communicate effectively using data.
Curriculum Learning Objectives

Demonstrate knowledge of...

1. foundational mathematical concepts relevant to data analysis
2. basic principles in computer science
3. foundational statistical concepts
4. basic principles in statistical computing
Curriculum Learning Objectives

Demonstrate the ability to...

5. take a real-world data analysis problem, formulate a conceptual approach to the problem, match aspects of the problem to previously learned theoretical and methodological tools, break down the solution into a step-by-step approach, and implement a working solution in a modern software language

6. communicate effectively in data analysis projects
ASA Guidelines: Key Points

• Increased importance of data science – the ability to “think with data”
  • New course in exploratory data analysis and statistical computing
  • Our entire Data Science curriculum was built with this point in mind

• Real applications as a major component available early in the curriculum
  • First-year seminar in data science
  • Senior year two-quarter capstone project course; team taught by Statistics and CS faculty
ASA Guidelines: Key Points

- More diverse models and approaches
  - Bridging both computer science and statistics courses emphasizes both statistical and algorithmic models
- Ability to communicate
  - Many required courses already require student presentations, group projects, and written reports
  - Senior year two-quarter capstone project course
ASA Guidelines: Skills

- Statistical methods and theory
  - Core Statistics courses

- Data manipulation and computation
  - Core computer science courses
ASA Guidelines: Skills

- **Mathematical foundations**
  - Calculus (through multivariate), linear algebra

- **Statistical practice**
  - ICS writing course
  - Senior year two-quarter capstone project course
  - Projects, presentations, reports included in many courses

- **Discipline-specific knowledge**
  - Some experience through senior capstone course
  - …just not enough room in the major!
Ideas that were not implemented

• Requiring a field of application (e.g., 2-3 courses in a discipline that uses statistics)

• Internships in Data Science with local companies

• “Tracks” or “concentrations” within the Data Science major
Data Science vs. Statistics

• B.S. in Statistics planned for future
  • Less computer science, more statistics
    • Possible future electives such as experimental design, time series analysis, etc.
  • Better preparation for graduate study in Statistics
    • More mathematics/theory (e.g., real analysis)
    • Include discipline-specific knowledge (e.g., 2-3 courses in applied area)
• B.S. in Data Science seen as more applied, possibly a terminal degree
Challenges and Next Steps

• Fast-changing data science technologies
• Large demand for the major → larger class sizes (currently 220 in each intro stat course and 400 in each intro CS course)
  • How to provide frequent and effective feedback? opportunities for effective communication (writing and presenting)?
• Undergraduate minor in Data Science
• Review and assessment of the success of the major
  • How do we know students are learning what we want them to learn?
  • Do we need to add material or modify any courses in the major?
Thank you!

Stacey Hancock
University of California, Irvine
stacey.hancock@uci.edu
References

