Revamping the Undergraduate Curriculum in Light of the ASA Guidelines

Kari Lock Morgan
Department of Statistics
Penn State University

klm47@psu.edu

JSM 2015
ASA Guidelines

• ASA Curriculum Guidelines (2014)

• Beth Chance, Steve Cohen, Scott Grimshaw, Jo Hardin, Tim Hesterberg, Roger Hoerl, Nick Horton (chair), Chris Malone, Rebecca Nichols, Deb Nolan

• Key points:
  – Increased importance of data science
  – Real applications
  – 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
Penn State Statistics Major

• Applied data analysis requirements:
  – intro stat, regression, ANOVA, SAS, capstone
• Math/probability/theory requirements:
  – calc I, II, III, matrices, probability, math stat
• 4 upper level stat (or math) electives
• Intro to computer science
• Gen-ed courses
• Option-specific requirements
Stat Major Options

6-7 courses in another field:

• Applied: minor in another field (not math)
• Actuarial: econ, finance, risk management
• Biostatistics: biology courses
• Computing: more computer science
• Graduate studies: more math
Process for Revamp

1. Map topics mentioned in ASA Guidelines to current curriculum; what’s missing?
2. Conversations with capstone instructor
3. Feedback from students
4. Discuss and decide on desired changes
5. Implement desired changes within the existing framework (hard!)
<table>
<thead>
<tr>
<th>Course(s) in which topic is covered</th>
<th>Current</th>
<th>New</th>
</tr>
</thead>
<tbody>
<tr>
<td>Design studies</td>
<td>461, 466</td>
<td>461, 466</td>
</tr>
<tr>
<td>Use graphical and other means to explore data</td>
<td>200, 462, 463, 470</td>
<td>200, 270, 461, 464, 470</td>
</tr>
<tr>
<td>Build and assess statistical models</td>
<td>462, 463, 470</td>
<td>462, 463, 470</td>
</tr>
<tr>
<td>Employ a variety of formal inference procedures (including methods)</td>
<td>200, 461, 464, 440</td>
<td>200, 270, 461, 464</td>
</tr>
<tr>
<td>Draw appropriate conclusions</td>
<td>lots</td>
<td>lots</td>
</tr>
<tr>
<td>Foundation in theoretical statistics principles for sound analyses</td>
<td>414, 415</td>
<td>414, 415</td>
</tr>
</tbody>
</table>

**Statistical Theory**

- Distributions of random variables                                    | 414     | 414     |
- Likelihood theory                                                     | 415     | 415     |
- Point and interval estimation                                         | 200, 415 | 200, 270, 415 |
- Hypothesis testing                                                    | 200, 415, 464 | 200, 270, 415, 464 |
- Decision theory                                                       | 470     | 415?    |
- Bayesian methods                                                      | 470     | 415?    |
- Resampling methods (bootstrapping and permutation tests)              | 440     | 270, 440|

**Exploratory Data Analysis**

- Visualization (including advanced)                                    | 200     | 200, 270, 470? |
- Visualization early for errors and anomalies                          | 200, 462 | 200, 462, 470 |
- Smoothing/kernel estimation                                           | 464 or 440? | 464 or 440? |
- Spatial methods                                                       | ???     | ???     |
- Mapping                                                               | ???     | ???     |

**Design of Studies**

- Data collection                                                       | 200, 461, 466 | 200, 270, 461, 466 |
- Random assignment                                                     | 200, 461 | 200, 461 |
- Blocking and stratification                                           | 461     | 461     |
- Adaptive designs                                                      | 461     | 461     |
- Efficiency (power?)                                                   | 461? 466? | 461, 466 |
<table>
<thead>
<tr>
<th>DATA MANIPULATION AND COMPUTATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facile with professional statistical software and other appropriate tools for data exploration, cleaning, validation, analysis, and communication</td>
</tr>
<tr>
<td>Able to program in a higher level language (write functions, utilize control flow in a variety of languages and tools such as Python, R, SAS, or Stata)</td>
</tr>
<tr>
<td>Think algorithmically</td>
</tr>
<tr>
<td><strong>Use simulation-based statistical techniques</strong></td>
</tr>
<tr>
<td>Undertake simulation studies</td>
</tr>
<tr>
<td>Manage and manipulate data, including joining data from different sources and formats and restructure data into a form suitable for analysis</td>
</tr>
<tr>
<td>Well-documented and reproducible way</td>
</tr>
<tr>
<td><strong>Software and tools</strong></td>
</tr>
<tr>
<td>Use of multiple data tools</td>
</tr>
<tr>
<td><strong>Accessing and Manipulating Data</strong></td>
</tr>
<tr>
<td>Judge data quality</td>
</tr>
<tr>
<td>Methods for addressing missing data</td>
</tr>
<tr>
<td>Work with <strong>csv</strong></td>
</tr>
<tr>
<td>Work with <strong>JSON</strong> (javascript object notation)</td>
</tr>
<tr>
<td>Work with <strong>XML</strong></td>
</tr>
<tr>
<td>Work with databases, database systems</td>
</tr>
<tr>
<td>Work with text data</td>
</tr>
<tr>
<td>Well-documented and reproducible</td>
</tr>
<tr>
<td>Breaking down a problem into modular pieces</td>
</tr>
</tbody>
</table>
1. We need data science!

- First action item: propose a new course in data science
  "Data Science Through Statistical Reasoning and Computation"
- Great for our students, but…
  
  WHO’S GOING TO TEACH IT???

- Challenge #1: Our faculty don’t know the data science topics we want our undergrads to know
Side note: More Data Science?

• Proposed data science major!
• Inter-college, joint between
  – College of Science (statistics, math, ...)
  – College of Engineering (computer science...)
  – College of Information Sciences and Technology

Increasing importance of data science!
2. Better Applied Courses

• Action item #2: make applied courses more rigorous

• Problem 1: Intro stat is the only prerequisite for most; lots of review

• Problem 2: Lots of non-majors; large courses

• Problem 3: Almost no sequential ordering

• Problem 4: No math prerequisites, but math is used
2. Better Applied Courses

• Action item #2: make applied courses more rigorous

• Problem 1: Intro stat is the only prerequisite for most; lots of review
  – Better intro?
  – STAT 2 as a prerequisite for all others?

• Problem 2: Lots of non-majors; large courses
  – Additional applied course for majors only?
  – Separate courses for majors and non-majors?
  – STAT 2 for everyone, upper-level mostly majors?

• Problem 3: Almost no sequential ordering
  – Order courses?

• Problem 4: No math prerequisites, but math is used
  – Add math prerequisites?
  – Minimize necessary math?
Applied Sequence Options

1. Intro -> Most applied courses
2. Intro -> STAT 2 -> More applied courses
3. Intro -> Applied courses for non-majors
   Intro -> Applied courses for majors
4. Intro -> Applied for all -> Extra applied for majors only (pre-capstone)

Challenge #2: Strong program for majors while keeping other disciplines happy.

Real applications, more diverse models and approaches
Challenge #2: Strong program for majors while keeping other disciplines happy.

Current Applied Courses:

1. Intro Stat → Applied 1 → Applied 2 → Capstone
2. Intro Stat → STAT 2 → Applied 2 → Applied 3 → Capstone
3. Intro Stat → Regression → ANOVA → Applied 3 → Capstone
4. Intro Stat → Regression → ANOVA → Capstone

Real applications
More diverse models and approaches
3. Better synthesis

• “Programs are encouraged to be creative with their curriculum to provide a synthesis of theory, methods, computation, and applications.”

• Where/how?
  – Electives?
  – More order/sequence of classes?
  – Additional class?
3. We’re doing a lot right!

• Capstone course that focuses on communication, teamwork, and real applications (mock consulting)

• Options provide discipline-specific knowledge and flexibility

• Familiarity with Minitab, SAS, R, C++

• DataFest provides big data, real application, teamwork, synthesis of computing, methods, communication
Key Points from Guidelines

• Increased importance of data science
  – new data science course!

• Real applications
  – good, but add more

• More diverse models and approaches
  – rework applied courses to improve

• Ability to communicate
  – stressed in capstone course