Teaching an Introductory Statistics Course: Challenges and Strategies

Kosuke Imai

Joint work with Jonathan Olmsted

Department of Politics
Princeton University

Curriculum Innovation Fund–Faculty Panel

November 18, 2014
Let’s Look at Some Data

- Non-politics introductory statistics courses in social sciences:
  - 5 year average: 2008/09–2013/14
  - ECO 302, PSY 251, SOC 301, WWS 200, WWS 332

<table>
<thead>
<tr>
<th></th>
<th>Lectures</th>
<th>Assignments</th>
<th>Readings</th>
<th>Precepts</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistics</td>
<td>3.2</td>
<td>3.3</td>
<td>3.1</td>
<td>3.6</td>
<td>3.1</td>
</tr>
<tr>
<td>All PU courses</td>
<td>3.8</td>
<td>3.7</td>
<td>3.7</td>
<td>4.0</td>
<td>3.9</td>
</tr>
</tbody>
</table>

- Politics introductory statistics courses (last time taught by us):

<table>
<thead>
<tr>
<th></th>
<th>Lectures</th>
<th>Assignments</th>
<th>Readings</th>
<th>Precepts</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>POL 345 (2011)</td>
<td>4.0</td>
<td>3.8</td>
<td>3.7</td>
<td>4.2</td>
<td>4.1</td>
</tr>
<tr>
<td>POL 245 (2014)</td>
<td>4.4</td>
<td>3.9</td>
<td>3.5</td>
<td>3.9</td>
<td>4.3</td>
</tr>
</tbody>
</table>
Students are not interested in statistics:

<table>
<thead>
<tr>
<th></th>
<th>Professor</th>
<th>Distribution Requirement</th>
<th>Departmental</th>
<th>Certificate Program</th>
<th>General Interest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistics</td>
<td>0%</td>
<td>20%</td>
<td>71%</td>
<td>3%</td>
<td>6%</td>
</tr>
<tr>
<td>All PU courses</td>
<td>6%</td>
<td>12%</td>
<td>32%</td>
<td>7%</td>
<td>42%</td>
</tr>
</tbody>
</table>

“Professor Imai tried hard to make statistics interesting. But, statistics is boring.”

Students have weak mathematical and programming background

“as a person not naturally inclined towards statistics and probability, I don’t feel at all qualified to pass judgement on how the course might have been improved.”
Two Strategies

1. **Motivating students**
   - Statistics as a necessary tool for modern social science research
     - Junior papers and senior thesis
     - Reanalysis of data from published research
   - Statistics as a useful skill for post-graduate career
     - Stories from course alumni/alumnus in various industries
     - Use of real-world examples

2. **Helping students learn**
   - Short but frequent assignments
     - Pre-precept assignments
     - Problem sets, quizzes
   - Hands-on instruction in computer labs
     - Detailed handouts
     - Practice questions
   - Outside-of-classroom assistance
     - Extra office hours, McGraw study halls
     - Piazza online discussion board
POL 245: Visualizing Data

- A new course offered first in Summer 2013
- Instructor: Jonathan Olmsted, Course head: Kosuke Imai
- Supported by the 250th Anniversary Fund
- Key idea: Teach data analysis before statistics
- POL 345 teaches both statistical concepts and data analysis skills
- This is too much for some students
- Take the first half of POL 345 and expand it into a full course
  - analyze a variety of structured and unstructured data
  - show data analysis and statistics are fun and relevant
- Challenge: Can we teach this course well so that students are encouraged to take the next statistics course?
Freshman Scholars Institute

- 6-week long summer school for a subset of incoming freshman
- an incubator for new courses

- Enrollment: 30 students
  - come from “disadvantaged” background
  - the first to go to college in their family
  - lack mathematical and computing background

- Goals:
  - transition them from high school to college
  - get them used to Princeton before the semester starts
  - offer head start by earning early PU course credits

- Similar programs at other schools: http://nyti.ms/1gjJOoU

- A hard test for our teaching strategies
Overview of POL 245

• Modules contents:
  1. Identifying causal effects (racial discrimination)
  2. Discovering patterns (political polarization)
  3. Making predictions (election forecasting)
  4. Textual data (federalist papers)
  5. Network data (supreme court citation)
  6. Geospatial data (Walmart expansion)

• Module format:
  1. Two 50 minute lectures
  2. Two 80 minute lab sessions
  3. One 80 minute guest lecture from industry, discussion over lunch (NYT, Facebook, Political and Energy consulting firms)

• Assignments:
  1. Four problem sets with no collaboration
  2. Six short non-graded pre-lab assignments
  3. One take-home midterm
  4. One collaborative final project and group presentation
Results and Next Steps

• Some encouraging feedback:

“The course was a lot of fun and really interesting and I plan on taking the next level of the course.”

“I really enjoyed this course, and that is why I want to take POL 345.”

“I felt it gave me a very true sense of what to expect at Princeton.”

• Next steps:
  1. Write a book to improve the course and impact beyond Princeton
  2. Completely flip the course
  3. Offer the course during the regular semester
  4. Develop a curriculum for a sequence of statistics and machine learning courses as part of the certificate program