Building and Training the Next Generation of Survey Methodologists and Researchers

Rebecca Nugent

Department of Statistics
Carnegie Mellon University

NCRN Spring Meeting
May 8, 2015
NSF-NCRN: The Carnegie Mellon Node

Members:
- PIs: Fienberg, Eddy
- Co-PIs: Acquisti, Nugent
- Faculty: Murray, Steorts (moving to Duke)
- Post-docs: Brandimarte (moving to University of Arizona)
- Students: Charest (Universite Laval), Cuellar, Gallagher, Hall (Etsy, Inc), Kim, Kurtz (PNC), Pane (dunnhumby), Popov (UPMC Western Psych), Sadinle (Duke), Ventura, and undergrads

Research Topics:
- Capture-recapture
- Censuses and Surveys
- Confidentiality, Privacy, and Online Self-Disclosure
- Record Linkage
- Small Area Estimation

http://www.stat.cmu.edu/NCRN
Educational Program

Goals:

- Increase number of students who are well-versed in statistical issues facing U.S. Census
- Increase number of students able to develop and implement appropriate analyses
- Support research projects (both summer and academic year)
- Incorporate research and related statistical problems in both undergraduate and graduate classes
- Create materials/modules that are reproducible

Framework/Expectations:

- Vertical Integration: senior faculty, junior faculty, graduate students, undergraduate students
- Regular research or class meetings
- Presentations, Reports, Posters
- Interdisciplinary work/audience
Carnegie Mellon and Graduate Statistics

- **Master’s in Statistical Practice Program**
  - One year program focusing on professional industry preparation
  - In-depth methods, practical applications, consulting projects
  - about 25 students a year

- **PhD program**
  - Statistics, joint programs with Heinz Public Policy, PIER, Machine Learning, CNBC, Engineering & Public Policy
  - Connections all over campus (research, educational programs)
  - Strong focus on interdisciplinary research; working at interface
  - Start research early (e.g. Advanced Data Analysis project)
  - about 8-12 students a year

All graduate classes oversubscribed by programs all over campus
Carnegie Mellon and Undergraduate Statistics

- ≈ 6500 undergrads at Carnegie Mellon University
- Statistics in Dietrich College of Humanities & Social Sciences
- Dietrich College ≈ 1000 UGs
- At least one intro. stats. course (usually as freshmen)
- ≈ 1400 – 1500 undergrads in stat classes per semester
- ≈ 200 majors
  - ≈ 60 primary Statistics majors
  - ≈ 95 in Economics-Statistics major
  - ≈ 40 in Statistics-ML major
- 87 students with additional major in Stats or Econ-Stats (MathOR, Business, Social/Decision Sciences, Psych, Creative Writing, Music, Philosophy... )
Overview of Program

▷ Solid theory background; thorough exposure to methodology
▷ Lower level electives: graphics & visualization, sample survey methods, statistical computing
▷ Advanced electives: undergraduate research, stochastic processes, statistical learning, multivariate analysis, multilevel/hierarchical models, data mining, methods in eli
▷ Self-defined “capstones”, prep for grad school and industry
   Offer options; students self-select (can help them)
▷ Motivated by large, real research problems - not textbooks
▷ Everyone takes same methods sequence;
   focus on scientific writing, interdisciplinary work
▷ **Industry:** tend to focus on methods courses, project
   competitions, professional development activities
▷ **Grad School:** tend to take research class, senior honors thesis, RA positions; *also* take extra methods courses,
   try to get them to take more math
Existing Classes and Projects

- Graduate level Sampling/Survey course (Fienberg, reoccurring)
- Discrete Multivariate Analysis (Fienberg, Sadinle-Garcia)
- Census-related data projects (Fienberg, Eddy, Nugent, others)
  - Modern Regression
  - Advanced Data Analysis
  - Special Topics courses
- *Sample, Survey, & Society* (Fienberg, Eddy, Nugent, Murray)
  - about 50 students; broadest range of majors of all stat classes
  - is a General Education class option for Dietrich college
  - material on sampling, survey design and analysis
  - students design, implement, and analyze a survey project
  - able to incorporate U.S. Census-related topics easily (had Renee Ellis give a guest lecture!)
New/Related Classes

- **2010 Census Freshmen Seminar**
  - Fienberg (reoccurring)
  - capped at 17; small class experience
  - Census-related topics
  - several statistics majors; at least two to grad school

- **Data Matching Methods and Their Uses**
  - Fienberg, Nugent, Ventura; Fall 2013
  - Master’s, undergraduate
  - Lectures, HW, Record Linkage project
  - 12 students: three to PhD programs, two to master’s programs

- **Intro to Privacy and Confidentiality**
  - Steorts; Fall 2014
  - Workshop for master’s level students

Upcoming: *Data Privacy*, Fienberg, Fall 2015
revamping *Sample, Survey, & Society*, Murray, Spring 2016
New Modules/Projects

- Statistical Graphics & Visualization
  - lower level elective class for students all over campus
  - extremely oversubscribed; long waitlists
  - extract, summarize, visualize features for low- and high-dimensional data
  - revamped to include group project on visualizing population characteristics using U.S. Census data

- Undergraduate Research
  - upper level elective; small, invitation-only course
  - students are paired with faculty clients
  - semester-long research project
  - Record Linkage (Steorts, 2014); Disease Spread/Synthetic Populations (Eddy, 2015)

Upcoming: Sample, Survey, Society; working with Heinz school on their sampling courses
Research Program (outside of thesis work)

- PhD Advanced Data Analysis projects
  - Data Swapping
  - Capture-recapture
  - Record Linkage
  - Synthetic Populations

- Master’s projects (occasional)

- Undergraduate Research projects
  - Summer Research Program (few students each year)
  - Research Assistantship positions (each semester)
  - Senior Honors Thesis
What are we doing well? What could we improve?

- **Successes**
  - Hundreds of students exposed to U.S. Census-related projects from wide range of disciplines (statistics, social sciences, etc)
  - Working with researchers and educators across campus and other universities to develop materials and programs
  - Active research agenda from freshmen to PhD level

- **Areas for Improvement/Future Plans**
  - Continue to modernize courses and modules
  - More active dissemination of materials
  - Continue work with Public Policy/Social Science MS programs
  - Better connections for students to statistical agencies
  - Facilitate more active recruiting; bring more U.S. Census people to us

http://www.stat.cmu.edu/NCRN