BRIEF PROJECT DESCRIPTIONS
36-490: Undergraduate Research, Spring 2010
Brian Junker & Cosma Shalizi

Using DocuScope to Grade English Writing in China

Contact: David Kaufer (Dept. of English, CMU, kaufer@andrew.cmu.edu) or Suguru Ishizaki (Dept. of English, CMU, suguru@cmu.edu)

There are many more people in mainland China who want to learn English than there are teachers to teach and grade their work. Especially difficult and time-consuming is grading essays and other written work in English. During his Fall 2009 sabbatical in China, Prof. Kaufer started using DocuScope (http://betterwriting.net/projects/fed01/dsc_fed01.html) to summarize and visualize features of English text written by Chinese students. Dr's Kaufer and Ishizaki wish to find the best statistical analyses of this data, to help teachers quickly provide feedback and instruction to large classes.

What Teaching Activities Help Students More?

Contact: Lindsay Clare Matsumura (Learning Policy Center, University of Pittsburgh, lclare@pitt.edu)

Teachers can do many things to help students become better readers. Do class discussions of reading help the most? Do homework assignments help the most? In this study, almost 100 teachers were graded on the quality of the in-class discussions they led, and on the quality of the homework they assigned to students. Students' scores on end-of-year reading comprehension tests were also collected, as well as other information about the students, schools and teachers. What factors contribute most to students' reading comprehension?

CEO Philosophy and Stakeholder Orientation of Companies

Contact: Andrea K. Young, Strategic Management and Public Policy, George Washington University, alk.young@comcast.net

Firms' top excecutives may hold different predominant philosophies, such as: *Pragmatic*, desiring to maximize output relative to input, with measurable ratios and results; *Intellectual*, a "sensemaking" view that tends to value activity or effort if it is consisted with a set of established guidelines or assumptions; *Humanist*, valuing enduring personal relationships; and *Organization*, identifying with, promoting, and wishing to remain with a particular firm. At the same time, the firm's strategic decisions may consider stakeholders (employees, customers, suppliers, environments, communities, and shareholders) in different ways, e.g.: *Narrow*, consideration of a single stakeholder, or *Broad*, consideration of many stakeholders. In this project, survey data from hundreds of respondents at various firms will be used to find connections between CEO philosophy and stakeholder orientation of strategic decision-making.

Classifying Patients' Need For And Use Of Pain Relief Medication

Contact: Alon Ben-Ari (Department of Anesthesiology, University of Pittsburgh, alon.benari@gmail.com)

A PCA (patient-controlled analgesia) pump is an electronic pump that delivers a prescribed amount of analgesic intravenously when the patient activates a button. Patients can receive medicine when they need it, instead of having to wait for nursing staff. In addition, the pump reduces the chances for medication errors, since it is programmed per the physician's order for amount and interval between doses and "locks out" the patient if he or she attempts to self-administer too often (http://en.wikipedia.org/wiki/Patient-controlled_analgesia). But different patients tend to use PCA pumps differently. In this project we will examine records of patients' minute-by-minute use of a PCA pump over approximately 16 hours, to try to classify patients according to their (unobservable) need for and (observable) use of pain-killing drugs. The initial data set consists of 20 patients, but additional data can be obtained if it is useful.

Problems in Statistical Astronomy

Contact: Peter Freeman (Dept. of Statistics, CMU, pfreeman@cmu.edu) or Joey Richards (Dept. of Statistics, CMU, jwrichar@stat.cmu.edu)

The International Computational Astrostatistics (InCA) Group is comprised of researchers from Carnegie Mellon University and University of Pittsburgh, as well as other locations around the globe. The goals of our group are to develop and apply novel statistical methods to inference problems in astronomy and cosmology, with an emphasis on computational nonparametric approaches. Here are three interesting undergraduate research projects.

- Star Formation History (Richards)

  Star formation history (SFH) parameters (age, mass, metallicity) for a sample of galaxies using their observed spectra from the Sloan Digital Sky Survey. However, because of the small aperture of the spectral fiber, the spectra sub-sample the light from each galaxy, and thus our estimates suffer from what the astronomers coin "aperture bias". The main problem here is to estimate the SFH parameters of galaxy light as a function of its photometric color (using the spectral estimates as a guide) and then correct our SFH estimates for each galaxy for the aperture bias.

- Source Detection and Characterization (Freeman)

  The data for this project consists of two-dimensional binned images, taken with charge-coupled devices (CCDs) that are sensitive to X-rays. (X-rays have wavelengths shorter than those the eyes can see; a typical wavelength is 1 Ångstrom, or 10^{-10} meters.) The object of the project is to develop algorithms for doing multiple comparisons, i.e., sequential hypothesis tests over all pixels, to provide a list of interesting X-ray sources (locations, fluxes, etc.) in the images.

- lare Detection and Characterization (Freeman)

  The data for this project consists of lists of photon arrival times. The object of the project is to develop an algorithm for detecting flares within these data, for instance by determining when the rate of photon arrivals changes significantly within a short time period. You'll apply this algorithm to many observations, and you'll try to answer the following questions: how many flares are there in each dataset? What is the significance of each detected flare? What are the flare characteristics (starting time, length, strength)? What confidence intervals can you place on these characteristics? How sensitive is your method, i.e., given a flare of strength S that has some defined shape as a function of time, how often does your algorithm detect it?