Strengthening Bridges between Statistics and the Natural Sciences

Designated workshop liaison: Jo Hardin, Pomona College, jo.hardin@pomona.edu

Liberal arts colleges have been the source and test bed for reform of introductory statistics for more than a decade. These changes have been internal within the statistics community, motivated by concerns about the effectiveness of pedagogy. The results have been much greater emphasis on statistical literacy, use of real and more complex data, less emphasis on abstract probability, and additional use of computer technology. Modern textbooks, which are now used nationwide, reflect the changes.

Despite the successes in improving statistics education to reflect the internal values of the field, the changes do not always address directly the evolving needs of other disciplines. There has been an increasing sophistication of the use of statistics in fields as diverse as political science, economics, neuroscience, psychology, sociology, biology, and environmental science.   In molecular biology, for example, technology has made possible the collection of huge data sets which have brought new statistical analysis challenges. Researchers are now tackling far more complex problems and using sophisticated methodologies that have only recently started to be incorporated into the undergraduate statistics curricula. The methodological changes, absent support from the statistics curriculum, have constrained the involvement of undergraduate students in real-world research projects.

In addition to the increasing complexity of methods, the last two decades have seen a dramatic increase in recognition of the importance of quantitative arguments. A number of national reports have called for change in the quantitative education of science students (e.g., the Bio2010 and MAA CRAFTY reports). This year's recommendations from HHMI-AAMC on the scientific preparation of pre-medical students "release[s] the student from specific course requirements, but ... require[s] each institution to identify the instructional means by which the necessary competencies can be gained." In regard to quantitative matters, the necessary competencies are very different from those gained from the traditional year of calculus. They include the ability to:

- articulate scientific questions and hypotheses
- read and analyze results from a paper in the scientific literature
- undertake and critique scientific inquiry
- apply the principles and approaches of statistics and epidemiology
- describe, analyze and make inferences using mathematical and statistical models
- extract relevant information from large data sets

It's clear that statistics is central to the quantitative education of scientists. In order to be able to fulfill their proper role, statistics curricula will have to engage the specific needs of today's science disciplines. It's doubtful that the conventional curriculum, however sophisticated the pedagogy, addresses the above list in any but the most abstract way. The problem is not that statisticians are unaware of the needs of the sciences --- indeed many statisticians work in intensive collaborations across diverse areas of science. Rather statistics, like many academic disciplines, has evolved according to internal standards and, at most liberal arts colleges, statistics is more closely aligned with mathematics than science.

This workshop seeks to reinforce the bridges between statistics and the natural sciences, enhancing the scientifically relevant practice of applied statistics into the curriculum. Doing so will help realign the introductory and intermediate statistics curricula with science curricula to provide a solid foundation for quantitative education and to ensure that the next generation of researchers have appropriate quantitative capacities. The primary focus will be on the applied statistics curricula, with the primary audience statisticians from the Mellon 23 institutions, augmented by a number of scientist colleagues from their institutions. Hosting the workshop at the Claremont Colleges facilitates participation from all of the local colleges, in particular, by inviting natural scientists from Pomona, Harvey Mudd and Scripps to participate in break-out sessions designed specifically to address needs of a specific discipline.

The workshop will be hosted by Pomona in January 2011. Jo Hardin (Pomona), Nicholas Horton (Smith) and Danny Kaplan (Macalester) will lead the workshop. Other faculty, primarily in statistics, have agreed to serve as members of the organizing committee, as well as attend the workshop. We have the opportunity to involve many local faculty within the sciences, computing and mathematics from the Claremont Colleges.

The goals of the workshop include review of the introductory and intermediate statistics curricula to foster greater connection with efforts underway in the natural sciences. The participants will brainstorm and refine other curricular (partial credit, J-term or summer programs) and co-curricular (journal clubs, statistical consulting structures, mentored research programs) programs that are appropriate within a liberal arts setting. In addition to the full report and assessment of the workshop, a paper describing the results and necessary next steps for implementation will be prepared.

Workshop Leaders

Jo Hardin, Department of Mathematics, Pomona College
(http://pages.pomona.edu/~jsh04747) Professor Hardin received her PhD from the University of California, Davis in 2000. Dr. Hardin was the recipient of the American Statistical Association Waller Award in 2007 in recognition of her excellence and innovation in teaching introductory statistics. She has published in *Bioinformatics, Computational Statistics and Data Analysis, The Journal of Computational and Graphical Statistics, Statistical Applications in Genetics and Molecular Biology, STATS, Chance,* and *Functional Ecology*. Dr. Hardin collaborates with biologists at Pomona College and Harvey Mudd College as well as medical researchers at Cancer Research and Biostatistics (in Seattle, WA.) She is a member of the American Statistical Association, the Institute of Mathematical Statistics, the Caucus for Women and Statistics, and Phi Beta Kappa. She serves as a representative-at-large for the Statistics Education Section of the American Statistical Association.

Nicholas Horton, Department of Mathematics and Statistics, Smith College
(http://www.math.smith.edu/~nhorton) Professor Horton received his doctorate from Harvard University in 1999. Dr. Horton was the recipient of the American Statistical Association Waller Award in 2009 in recognition of his excellence and innovation in teaching introductory statistics. His work has appeared in *Biometrics, Journal of the Royal Statistical Society-Series A, The American Statistician, The American Journal of Public Health, The New England Journal of Medicine* and other journals. Dr. Horton collaborates with investigators at Smith College (Biological Sciences, Psychology, Engineering), Harvard University, Boston Medical Center, Pavlov Medical University (St. Petersburg, Russia) and the University of Auckland. He is a member of the National Council of Teachers of Mathematics/American Statistical Association Joint Committee on Probability and Statistics, and the Executive Committee of the Statistics Education Section of the American Statistical Association.

Danny Kaplan, Department of Mathematics, Statistics, and Computer Science, Macalester College (http://www.macalester.edu/~kaplan) Professor Kaplan received his doctorate in biomedical physics from Harvard University in 1989. He holds an endowed chair and received Macalester's 2006 Excellence in Teaching award. His research work in electrophysiology has appeared in physics, physiology, cardiology and neuroscience journals. His textbooks (Understanding nonlinear dynamics, Statistical modeling: a fresh approach, and Introduction to scientific computation and programming) have been widely adopted. With support from HHMI and the Keck Foundation, he led Macalester's revision of its introductory calculus, statistics, and computation curriculum to serve better the needs of science students.

Workshop participants/planning committee members (who have confirmed that they will participate in helping to plan as well as attend the workshop)

Steve Adolf, Department of Biology, Harvey Mudd College.
Professor Adolph's research explores the ecology and physiology of lizards. He is currently working on improved statistical methods for estimating maximum animal performance (e.g., running and jumping) and correlations between different performance measures. He teaches ecology, introductory biology, biostatistics and mathematical biology. He is very interested in statistics education and recently published a teaching exercise in the Journal of Statistics Education.

Laura Chihara, Department of Mathematics, Carleton College.
Dr. Chihara taught previously at St. Olaf College and worked at MathSoft (makers of S-plus software) prior to moving to Carleton. She has been developing and strengthening the statistics offerings in the mathematics department and has been involved in the quantitative reasoning initiative as well as the spatial literacy initiative on campus.

Albyn Jones, Department of Mathematics, Reed College.
Dr. Jones' work has appeared in various journals and proceedings, including *The American Statistician, Chance, Neuron* and *Heredity*. He has collaborated in numerous academic fields, as well as medical, industrial and legal settings, and regularly consults with thesis students in the biological sciences.

Shonda Kuiper, Department of Mathematics, Grinnell College.
Professor Kuiper's work on creating interdisciplinary research projects for statistics courses has been supported by a three year National Science Foundation grant and published in The American Statistician and the Journal of Statistics Education. Dr. Kuiper's consulting work has led to publications in the International Journal for Parasitology, the American Midland Naturalist, and the American Institute of Chemical Engineers Journal. She is a member of the Consulting Section and the Statistics Education Section of the American Statistics Association. Dr. Kuiper currently leads the Isolated Statistician Group as well as the Post-Introductory Research Cluster for the Consortium for the Advancement of Undergraduate Statistics.

Shu-Min Liao, Department of Mathematics and Statistics, Amherst College.
Dr. Liao has been engaged in ways to improve statistics education in response to increased demand for quantitative training for natural science students. She has regularly attended statistical education workshops and is involved in collaborative research projects.

Matthew Neal, Department of Mathematics and Computer Science, Dennison University
Matthew Neal is a mathematician who has taught beginning, intermediate and upper level statistics classes for many years. He also serves as chair of the department. He is interested in developing beginning, intermediate, and advanced interdisciplinary statistical modeling courses that best serve both science students and mathematics majors.

Amy Wagaman, Department of Mathematics, Amherst College.
As a relatively new statistician at a liberal arts school with a developing statistics program and a section of intro stats with an environment focus, Dr. Wagaman is very interested in new developments integrating our statistics curricula more with the natural sciences. She has participated in the statistics education community through USCOTS and explored the use of new technology (personal response systems) at the University of Michigan.

Jeff Witmer, Department of Mathematics, Oberlin College.
Dr. Witmer developed an introductory biostatistics course at Oberlin after consulting with a number of the Oberlin science faculty; this course uses a book that he co-authored. During his years in the Dean's Office he promoted bioinformatics and a fledgling concentration in scientific modeling.

Strengthening Bridges between Statistics and the Natural Sciences

We propose to hold a 2.5 day workshop January 9-11, 2011 at Pomona College. The purpose of the workshop is to facilitate statistics curricular reform with respect to integrating the life sciences. Participants will join us for 2.5 days of intensive collaboration and will be selected from the member Mellon colleges.

January 9-11, 2011	Participants	Amount	Total
Stipends			
Organizers	3	$300	$ 900
Speakers	2	$200	$ 400
Travel			
Participants	15	$500	$7500
Guest Speakers	2	$500	$1000
Subsistence			
Participants (out-of-town)*	15	$450	$6750
Participant (local)**	3	$150	$ 450
Speakers*	2	$450	$ 900
Other			
Administrative Support	1	$1000	$1000
Photocopying	1	$200	$ 200
Grand Total			**$19,100**

* Hotel & food
** Food only

610 N. College Ave.
Department of Mathematics
Claremont, CA 91711
(909) 607-8717

jo.hardin@pomona.edu
424 Baughman Ave.
Claremont, CA 91711
(909) 624-5860

EDUCATION

Ph.D. in Statistics, University of California, Davis   2000
Dissertation: **Multivariate Outlier Detection and Robust Clustering
with Minimum Covariance Determinant Estimation and S-Estimation.**
Advisor: Dr. David M. Rocke.

M.S. Statistics, University of California, Davis   1997

B.A. Mathematics, Pomona College, Claremont, California   1995
Advisor: Dr. Donald Bentley.

EMPLOYMENT

Associate Professor of Mathematics, Pomona College   2008 - present

Assistant Professor of Mathematics, Pomona College   2002 - 2008

Staff Scientist, Fred Hutchinson Cancer Research Center   2000 - 2002

Lecturer, Seattle University, Albers School of Business and Economics   2001 - 2002

Associate Instructor, UC Davis, Division of Statistics   1997 - 1999

HONORS AND AWARDS

American Statistical Association Waller Education Award	2007
Appointed to CGU's Extended Graduate Faculty	2003
Project NExT Fellow	2002-2003
Graduate Student Association Travel Award, UC Davis	2000
Student Paper Competition Runner-Up, Joint Meeting of WNAR, IBS, & IMS	1999
Julius Blum Award, Division of Statistics, UC Davis	1995-1996
Phi Beta Kappa, Pomona College	1995
Sigma Xi, Pomona College	1995
Graduated Cum Laude, (B.A.), Pomona College	1995
Graduated with Honors, Mathematics Department, Pomona College	1995

SELECTED PUBLICATIONS AND PAPERS

- **Peer-reviewed articles**

Richards, J.[†], Hardin, J., Grosfils, E.; **Weighted Model-Based Clustering for Remote Sensing Image Analysis**, *Computational Geosciences*, 2009. (DOI: 10.1007/s10596-009-9136-z)

Hardin, J., Wilson, J.; **A note on oligonucleotide expression values not being normally distributed**, *Biostatistics*, **10**: 446-450; 2009.

Yiu, G.[†], McCord, A.[†], Wise, A.[†], Jindal, R.[†], Hardee, J.[†], Kuo, A.[†], Yuen Shimogawa, M.[†], Cahoon, L., Wu, M., Kloke, J., Hardin, J., Mays Hoopes, L.L.; **Pathways Change in Expression During Replicative Aging in *Saccharomyces cerevisiae***, *Journal of Gerontology, 63A*: 21-34; 2008.

Hardin, J., Mitani, A.[†], Hicks, L.[†], VanKoten, B.[†]; **A Robust Measure of Correlation between Two Genes on a Microarray**, *BMC Bioinformatics*, **8**:220; 2007.

Adolph, S., Hardin, J.; **Estimating Phenotypic Correlations: Correcting for Bias Due to Intraindividual Variability**, *Functional Ecology*, **21**:178-184; 2007.

Wise, A.[†], Hardin, J., Hoopes, L.; **Yeast Through the Ages: a statistical analysis of genetic changes in aging yeast**, *Chance*, **19**: 39-44; 2006.

Hardin, J., Hoopes, L., Murphy, R.[†]; **Analyzing DNA Microarrays with Undergraduate Statisticians**, *Proceedings of the Seventh International Conference on Teaching Statistics*, 2006.
http://www.stat.auckland.ac.nz/~iase/publications.php?show=17
course modules available at: http://pages.pomona.edu/~jsh04747/courses/microarray/modules.htm

Hardin, J., Rocke, D.; **The Distribution of Robust Distances**; *Journal of Computational and Graphical Statistics*, **14**: 928-946; 2005.

Hardin, J., Waddell, M., Page, D., Zhan, F., Barlogie, B., Crowley, J., Shaughnessy, J.; **Evaluation of Multiple Models to Distinguish Closely Related Forms of Disease Using DNA Microarray Data: an Application to Multiple Myeloma**, *Statistical Applications in Genetics and Molecular Biology*, **3**: article 10; 2004.

Hardin, J., Rocke, D.; **Outlier Detection in the Multiple Cluster Setting Using the Minimum Covariance Determinant Estimator**, *Computational Statistics and Data Analysis*, **44**: 625-638; 2004.

Durbin, B., Hardin, J., Hawkins, D., Rocke, D.; **A Variance-Stabilizing Transformation for Gene-Expression Microarray Data**; *Bioinformatics*, **18**: S105-S110; 2002.

Zhan, F., Hardin, J., Kordsmeier, B., Bumm, K., Zheng, M., Tian, E., Sanderson, R., Yang, Y., Wilson, C., Zangari, M., Anaissie, E., Morris, C., Muwalla, F., van Rhee, F., Fassas, A., Crowley, J., Tricot, G., Barlogie, B., Shaughnessy, J.; **Global Gene Expression Profiling of Multiple Myeloma, Monoclonal Gammopathy of Undetermined Significance, and Normal Bone Marrow Plasma Cells**; *Blood*, **99**: 1745-1757; 2002.

Coleman, D., Dong, X., Hardin, J., Rocke, D.M., Woodruff, D.L.; **Some Computational Issues in Cluster Analysis with no á priori Metric**; *Computational Statistics and Data Analysis*, **31**: 1-11; 1999.

[†]Indicates work done as an undergraduate.

Nicholas J. Horton

Education

Harvard University Biostatistics Sc.D., 1999 Harvard University Psychology A.B., 1987

Appointments

Associate Professor, Department of Mathematics and Statistics, Smith College, 2007–present

Adjunct Associate Professor, Department of Biostatistics, Boston University, 2007–present

Assistant Professor, Department of Mathematics and Statistics, Smith College, 2003–2007

Assistant Professor, Department of Biostatistics, Boston University School of Public Health; Department of Medicine, Section of General Internal Medicine, Boston University School of Medicine; Graduate School of Arts and Sciences, 2000–2003

Most Relevant Publications

Kleinman KP and **Horton NJ**. *SAS and R: Data management, statistical analysis and graphics*, Chapman and Hall/CRC Press, (2010).

Horton NJ, Roberts K, Ryan L, Suglia S, and Wright RJ. A maximum-likelihood latent variable model for multiple informants. *Statistics in Medicine*, (2008); 27:4992-5004.

Horton NJ and Kleinman KP. Much ado about nothing: A comparison of missing data methods and software to fit incomplete data regression models. *The American Statistician*, (2007); 61(1):79-90.

Horton NJ and Laird NM. Maximum likelihood analysis of logistic regression models with incomplete covariate data and auxiliary information. *Biometrics* (2001) 57:34-42.

Horton NJ and Lipsitz SR. Multiple imputation in practice: Comparison of software packages for regression models with missing variables. *The American Statistician*, (2001) 55(3):244-254.

Horton NJ and Laird NM. Maximum likelihood analysis of generalized linear models with missing covariates. *Statistical Methods in Medical Research* (1999) 8:37-50.

Other Significant Publications (* denotes undergraduate student)

Voss SE, **Horton NJ**, Tabucchi TH*, Folowosele FO* and Shera CA. Posture-induced changes in distortion-product otoacoustic emissions and the potential for noninvasive monitoring of changes in intracranial pressure, *Neurocritical Care*, 2006; 4(3):251-7.

Horton NJ and Switzer SS*. Statistical methods in the *Journal. New England Journal of Medicine*, 2005; 353(18):1977-1979.

Horton NJ and Fitzmaurice GM. Regression analysis of multiple source and multiple informant data from complex survey samples. *Statistics in Medicine*, (2004), 23(18):2911-2933.

Horton NJ, Brown ER, and Qian L*. Use of R as a toolbox for mathematical statistics exploration. *The American Statistician*, (2004) 58(4):343-357.

Horton NJ and Fitzmaurice GM. Maximum likelihood estimation of bivariate logistic models for incomplete responses with indicators of ignorable and non-ignorable missingness. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, (2002), 51(3):281-295.

Synergistic Activities

Former member, Scientific advisory committee, Reed College, Oregon Graduate Institute Superquest project (supercomputing institute for high school teams, 1992-1993)

Scientific reviewer for NIH / NSF (Standing member of Center for Scientific Review Community influences on health behavior study section)

Excellence in Teaching Award, Boston University School of Public Health, 2001

Smith College Student Government Excellent in Teaching Award, 2009

American Statistical Association Waller Award for innovations in teaching introductory statistics, 2009

Development of activities for the use of technology in the introductory, intermediate and mathematical statistics course.

Member of the American Statistical Assocation / National Council of Teachers of Mathematics Joint Committee on K-12 Probability and Statistics

Elected Vice Chair American Statistical Association Council of Chapters (District 2)

Elected Executive Committee member, American Statistical Association Section on Statistical Education

Collaborators and Co-editors

Garrett Fitzmaurice, Harvard University
Jo Hardin, Pomona College
Danny Kaplan, Macalester College
Ken Kleinman, Harvard University
Cliff Konold, University of Massachusetts/Amherst
Nan Laird, Harvard University
Michael Larsen, Iowa State University
Michael Lavine, University of Massachusetts/Amherst
Stuart Lipsitz, Harvard University
Sharon-Lise Normand, Harvard University
Richard Saitz, Boston University
Jeffrey Samet, Boston University
Susan Voss, Smith College
Chris Wild, University of Auckland
Byron Zamboanga, Smith College

Graduate Advisor

Nan Laird, Harvard University

Undergraduate thesis advisor for Mariel Finucane and Kristin Tyler, master's thesis committee member for Cynthia Chuang, Claudia DeYoung and Stefan Kertesz, Boston University, as well as doctoral thesis committee member for Yvette Cozier and Serkalem Demissie, Boston University, and Liam O'Brien and Heather Litman, Harvard University.

Biographical Sketch
DANIEL T. KAPLAN

Department of Mathematics, Statistics, & Computer Science   kaplan@macalester.edu
Macalester College   http://www.macalester.edu/~kaplan
Saint Paul, MN 55105   Voice/Fax: (651) 696-6599/6518

i. PROFESSIONAL PREPARATION

Swarthmore College, Swarthmore PA	Physics	B.A.	1981
Stanford University, Stanford CA	Engineering-Economic Systems	M.S.	1982
Harvard University, Cambridge MA	Biomedical Physics	M.S.	1986
Harvard University, Cambridge MA	Biomedical Physics	Ph.D.	1989

ii. APPOINTMENTS

9/05–	DeWitt Wallace Professor of Mathematics, Statistics, & Computer Science, Macalester College
1/08–7/08	Visiting Scholar, School of Biomedical Sciences, University of Queensland, Brisbane, Australia
9/00–8/05	Associate Professor, Department of Mathematics & Computer Science, Macalester College
9/96–8/00	Assistant Professor, Department of Mathematics & Computer Science, Macalester College
1/01–6/01	Academic Visitor, Mathematical Institute, University of Oxford, England
6/96–7/96	Enseignant Invité, Professeur Biologie, Université Paris 7, Denis Diderot
1/94–6/96	Assistant Professor, Department of Physiology, McGill University
1/91–1/94	Postdoctoral Fellow, Centre for Nonlinear Dynamics in Physiology and Medicine, McGill Univ.

iii. (a). FIVE PUBLICATIONS RELATED TO PROPOSED TOPIC

D. T. Kaplan (2009) *Statistical Modeling: A Fresh Approach.* www.macalester.edu/~kaplan/ISM

D. T. Kaplan (2007) "Computing and Introductory Statistics" *Technology Innovations in Statistics Education* 1(1):Article 5, repositories.cdlib.org/ulastats/cts/tise/iss1/art5

D. T. Kaplan (2003) *Introduction to Scientific Computation and Programming.* Brooks-Cole. www.macalester.edu/~kaplan/ScientificProgramming

D. T. Kaplan and L. Glass (1995) *Understanding Nonlinear Dynamics.* Springer-Verlag

D. T. Kaplan (2004) "Teaching computation to undergraduate scientists". Proceedings ACM Special Interest Group on Computer Science Education..

iii. (b). FIVE ADDITIONAL PUBLICATIONS

D. T. Kaplan (1999) *Resampling Statistics in MATLAB* Resampling Stats, Inc.

C. L. Bremer and **D. T. Kaplan** (2001) "Markov chain monte carlo estimation of nonlinear dynamics from time series" *Physica D*, 160(116-126)

P. Faure, **D. T. Kaplan**, and H. Korn (2000) "Synaptic efficacy and the transmission of complex firing patterns between neurons" *Journal of Neurophysiology* 84:3010-3025

B Pilgram and **D. T. Kaplan** (1999) "Nonstationarity and 1/f noise characteristics in heart rate" *American Journal of Physiology* 276, Regulatory, Integrative, and Comparative Physiology 45(1):R1-R9

D. T. Kaplan, JR Clay, T Manning, L Glass, MR Guevara, and A Shrier. "Subthreshold dynamics in periodically stimulated squid giant axons" *Physical Review Letters* 76(21):4074-4077

iv. SYNERGISTIC ACTIVITIES

1. PI, NSF Course, Curriculum, & Laboratory Improvement Phase II Project: Building a Community around Modeling, Statistics, Computation, and Calculus." NSF DUE 0920350, 2009-2012.
2. Steering Committee, Concentration in Community and Global Health. Epidemiology instructor. 2008-present
3. PI, Keck Data Fluency Project, Macalester College, 2004-2009
4. Co-organizer *Stat Chat*, a monthly meeting of local and regional statistics educators. 2004-present
5. Faculty Assessment Coordinator, Macalester College, 2005-2007
6. Founding Coordinator, Applied Mathematics and Statistics Major, Macalester College, 2007-present
7. Associate Director, Quantitative Methods for Public Policy Program, Macalester College, 2002-2005