NISS Parameters



Winter/Spring 2012

A quarterly newsletter from the National Institute of Statistical Sciences

NISS and Duke Receive Five-Year Grant from NSF for Triangle Census Research Network

NISS and Duke University have received a grant from the National Science Foundation (NSF) and the U.S. Census Bureau for the Triangle Census Research Network (TCRN). The award, one of eight nationwide under the NSF-Census Research Nodes program, is for nearly \$3 million and covers a five-year period. Jerome Reiter (PI), Mrs. Alexander Hehmeyer, Associate Professor of Statistical Science at Duke; and Alan Karr, director of NISS and co-PI, are collaborating on this project, along with other statisticians, economists and political scientists from Duke and NISS.

The grant will be used to improve how federal statistical ("FedStats") agencies disseminate data to the public and to researchers. Specifically, the TCRN will enhance FedStats agencies' capabilities by developing broadly-applicable methodologies in three interrelated areas: (i) disseminating public use data with high utility and acceptable disclosure risk, (ii) handling missing data and correcting faulty data in large complex surveys, and (iii) integrating information from multiple data sources. The TCRN will also offer educational opportunities to postdoctoral fellows, graduate students, and statisticians at federal agencies, helping to train future leaders in data dissemination research and practice.

The FedStats agencies collect data of all kinds that affect many people, including the decennial census, unemployment numbers and the Consumer Price Index. NISS has been and is in collaboration with many of these agencies, including (in addition to the Census Bureau) the Bureau of Labor Statistics, the Bureau of Transportation Statistics, the Energy Information Administration, the National Agricultural Statistics Service, the National Center for Education Statistics and the National

Center for Health Statistics. Among NISS' achievements are methods used nationally to produce high school graduation rates and crop forecasts, as well as a plethora of techniques and tools that support dissemination of high quality information derived from confidential data.

By building on these achievements as well as creating

new theory and methodology applicable to major Census Bureau data products, the TCRN's research will improve the hundreds of secondary analyses of these datasets. The interdisciplinary team of the TCRN, which consists of statisticians, economists, political scientists and operations researchers, will use these data products to answer questions in aging, economics, and social welfare that have important implications for policy making.

"The TCRN will improve the way we handle missing and faulty data by integrating paradigms from statistics and operations research," explained Karr, "The team will also



Photo by U.S. Census Bureau, Public Information Office

develop nonparametric Bayesian approaches for multiple imputation of missing data in high dimensions with longitudinal and multi-level aspects, as well as address central issues in data integration."

Affiliates Meeting in Washington DC

NISS-SAMSI Affiliates are invited to the paired events: a Workshop on Cell Suppression and the 2012 Affiliates Meeting with technical talks focusing on the upcoming SAMSI year-long programs: Data-Driven Decisions in Healthcare and Statistical and Computational Methodology for Massive Datasets. Programs for the workshop and the meeting will be published soon on the NISS and SAMSI websites. Each affiliate liaison is invited to attend and to bring a colleague whose interests intersect one of the three highlighted topics; junior colleagues are especially welcome.

May 14 (Monday)	8:30am – 3pm:	Cell Suppression Workshop
	3 – 5:30pm: 6pm:	Affiliates Meeting – NISS and SAMSI updates and business Affiliates Dinner
May 15(Tuesday)	8:30 – 2:30 pm: 9 – 12:15pm: 12:15 – 2:30pm:	Affiliates Meeting continued Invited talks on topics in Healthcare and Massive Data SAMSI - planning for 2013-2014, as well as Affiliates' roles in SAMSI

Separate registrations are required for the workshop and the affiliates meeting and program. Space in the workshop is limited but affiliates will have priority.

Topics in Healthcare and Massive Data: Agenda to be published on NISS and SAMSI websites. Talks will presage the SAMSI programs by addressing motivation, background material and technical issues for these 2012-2013 programs.

Cell Suppression Workshop

The purpose of the May 14 Cell Suppression Workshop is to expose government agencies and the Affiliate community to best practices and software for cell-suppression-based disclosure limitation, with emphasis on methods for statistical analysis of data containing suppressions, and alternatives to cell suppression. The Workshop is neither introductory nor highly technical. It is aimed at practitioners who have a general understanding of disclosure limitation, including consumers of disclosure limitation as well as those actively involved in research and development. It will focus on shared practical and methodological problems faced by statistical agencies.

In practice, cell suppress almost always means complementary cell suppression (CCS): in order to thwart accurate estimation of suppressed sensitive cell values,

additional, non-sensitive cells must also be suppressed. CCS has stimulated research and publications in statistics, mathematics, computer science and operations research since the 1970s, but without consensus about either how to do cell suppression or how to analyze data—especially magnitude data containing suppressions.

Participants in the workshop will work toward common understanding of the characteristics of and challenges posed by CCS, including its effects on data quality and usability, as well as what agencies releasing suppressed tabular data can do to facilitate and simplify analyses of them.

Attendance at this workshop is by invitation only; contact Lawrence Cox (cox@ *niss.org*), organizer of this workshop, if you are interested.



NISS Parameters is a quarterly publication of the National Institute of Statistical Sciences ©2012. All rights reserved. No part of this publication may be reproduced without the permission of NISS.

Postdoc Profile: Junheng Ma

Junheng Ma started working for NISS this past fall on the National Survey of College Graduates (NSCG) and National Survey of Recent College Graduates (NSRCG) as part of a project funded by the National Center for Science and Engineering Statistics at NSF. Last year, Junheng was a postdoc for SAMSI and was involved with the Analysis of Object Data program. He collaborated with people in the Statistical Inference for Functional Data working group.

Junheng grew up in the northern part of China in a mid-sized city called YangQuan. He went to college nearby and studied applied mathematics, then applied to graduate schools in the United States. He chose to go to Case Western Reserve University in Cleveland, Ohio where he studied statistics. Joe Sedransk was one of his two advisors while he was at Case. His thesis work, "Contributions To Numerical Formal Concept Analysis, Bayesian Predictive Inference And Sample Size Determination," involved survey methodology, which has been a good



a postdoc at SAMSI.

at NISS.

NISS Calendar of Events

Professor Murrad S. Taggu April 19-21, 2012 at UNC-Chapel Hill. ARA ELIGIBLE.

Network of Greater Georgia Institutions for Neuroimaging and Statistics (NOGGINS) Workshop April 20, 2012 at University of Georgia in Athens, GA. ARA ELIGIBLE.

NISS and SAMSI Affiliates Meeting and NISS Cell Suppression Workshop May 14-15, 2012 at Bureau of Labor Statistics (BLS) in Washington DC

Southern Regional Council on Statistics Summer Research Conference June 3-6, 2012 at Jekyll Island, GA. ARA ELIGIBLE.

8th International Purdue Symposium on Statistics "Diversity in the Statistical Sciences for the 21st Century" June 20-24, 2012 at Purdue University, West Lafayette, IN. ARA ELIGIBLE.

connection to the current work he is doing

Junheng said he is enjoying working on the project so far. He meets with Ivan

Carrillo-Garcia, another NISS postdoc, and Alan Karr, director of NISS, a couple of times a week to discuss what results they are finding so far. He said it has been an excellent

Jungheng Ma is a postdoctoral fellow at NISS this year. Previously, he was

environment to work in. He feels working on these surveys has helped to broaden his research activities.

When Junheng is not working at NISS he likes to spend time with his one and a half-yearold daughter, Sydney, and his wife, both whom are currently living in Boston.

Long-Range Dependence, Self-Similarity and Heavy Tails - International Conference in Honor of

Karr Elected into Membership of Johns Hopkins Society of **Scholars**

Alan Karr, Director of NISS, was elected into membership in the Johns established in May 1967 by the trustees of the University to honor distinguished former postdoctoral fellows, postdoctoral degree recipients, house staff and junior or visiting

faculty who have served at least a year at Johns Hopkins and thereafter gained marked Hopkins Society of Scholars. The Society was distinction elsewhere in their fields and for whom at least five years have elapsed since their last Johns Hopkins affiliation. Karr spent twenty years at Johns

Hopkins University in the Mathematical

Diebold Project Explores Failure Data for ATMs

Alan Karr and Stan Young are working on a project with Diebold Corporation. The principal focus of the project is to analyze machine- and human-generated failure data for automatic teller machines (ATMs). One thing the company hopes to achieve is to help predict when these failures may occur. Currently Karr and Young are looking at data from machines that are located indoors and that can only be used to withdraw money. Typically these machines are found in places like casinos and bars.

Young and Karr Propose Ways to Improve How **Observational Studies are Conducted**

Stan Young and Alan Karr published a non-technical article in the September issue of Significance magazine pointing out that medical and other observational studies often

produce results that are later shown to be incorrect, and—invoking a quality control perspective—suggest ways to fix the system.

Their central point is that the current system of publication in peer-reviewed journals relies on post-production inspection to ensure quality, a practice that has disappeared from modern industry in favor of controlling the process instead: quality control is now process control, not product control. They cite W. Edwards Deming, considered by many the most innovative thinker ever about quality, arguing not only for process control, but also that the problem lies with the managers—funders and journals—rather with than the workers—individual researchers who respond rationally to the current set of incentives.

Young and Karr describe both their and others' studies of the extent to which observational studies do not replicate. Published claims such as "coffee

causes pancreatic cancer," or "women eating breakfast cereal are more likely to have boy babies," have been refuted by subsequent studies and analyses. When these studies



reach the popular media and influence individual consumers, the burden falls not

Whiting School of Engineering from 1986-1992.

just on science but also on society. And

personnel, may have been squandered.

are asked, some will yield false positive

models until one is found that "fits the

data"). Publication bias is another issue:

papers reporting positive scientific results

(for example, an association between Type

A personalities and heart attacks) are more

likely to be published than those reporting

negative results, even though the latter may

a study is submitted for publication, the data

be split into two sets, a modeling data set

and a holdout data set. Journals would then

accept or reject papers based on the analysis

of the modeling data set without knowing

the results of applying the methods to the

results of the analysis of the holdout set.

the American Statistical Society.

holdout set. But then the journal would also

publish an addendum to the paper giving the

Significance magazine is published by

the Royal Statistical Society of the UK and

Young and Karr recommend that when

be as important scientifically.

even if there were no impact on the public,

scarce research resources, both money and

The paper describes several technical

difficulties with observational studies, among

them multiple testing (if enough questions

answers), bias (systematic error) and multiple

modeling (searching among mathematical

Sciences

He was a

Department.

Department

He served as

Chair of the

Department

from 1985-86.

He also served

Dean, G.W.C.

as Associate

Professor in the

from 1983-1993

various institutes.

Larry is currently serving as a Council member. The incoming ISI President, JC Lee, has identified Portfolio Projects for the Council; each project is assigned to two Council members. Larry serves on two projects: Professional Ethics and Accreditation.

Karr and Cox Teach Short Course on Data Confidentiality

NISS Outreach

Cox and Karr Speak at ISI Dublin

Alan Karr and Larry Cox taught a short course on the subject of confidentiality and disclosure avoidance in November at the National Center for Educational Statistics in Washington DC. Confidentiality is an area that is of great interest to many federal statistical agencies. 2012 Federal Committee on Statistical Methodology (FCSM) Research Conference, Jan. 10-12 2012

Alan Karr and Larry Cox presented papers during the "Simulation Models of Federal Surveys" session. Larry's paper, "The Case for Simulation Models of Federal Surveys" summarized the NISS Survey Simulation workshop. He talked about developing survey simulators and the issues of the scope of processes simulated. Allen's paper, "The World's Smallest Survey Microsimulator," described software to simulate surveys. It is a simple program based on assumptions designed to demonstrate the importance of simulating survey agents (interviewers and respondents) and their interactions.

ASA Statistics and Policy Workshop, Feb. 2-3, 2012

Larry Cox gave a talk at the workshop entitled, "A Simulation Laboratory for Federal Surveys." John Eltinge from the US Bureau of Labor Statistics was the discussant. Alan Karr presented "The World's Smallest Survey Microsimulator" talk at this workshop as well. NISS Hosts NC Science Fair Foundation

The North Carolina Science Fair Foundation (NCSFF) held its annual board meeting at NISS on December 9. NCSFF promotes science and engineering research by elementary, middle, and high school students, and partners with teachers to integrate scientific research into the classroom. NCSFF organizes science and engineering fairs to showcase and celebrate student research and learning. The board consists of teachers and professors from around North Carolina. Alan Karr greeted the group and told them about NISS and our mission and vision.

NISS held its 2011 annual meeting on November 4 and 5, at its headquarters in Research Triangle Park, NC. More than twenty members of the Board of Trustees and the NISS Corporation attended—a record. The focus of the meeting was breakout discussions on the NISS-SAMSI Affiliates Program and "Marketing NISS." Recognizing the role of the American Statistical Association (ASA) as one of NISS' ten parent organizations, the Board observed a moment of silence in remembrance of the late Martha Aliaga, former director of education at the ASA, whose memorial service was held on November 5.



Jim Rosenberger (Penn State U.), Xihong Lin (Harvard U.), Bob Rodriguez (SAS (Amgen), Fritz Scheuren (NORC at the U. of Chicago), Keith Soper (Merck).

Larry Cox and Alan Karr recently traveled to Dublin, Ireland to present at the 58th World Statistics Congress of the International Statistical Institute (ISI). Participants in the Congress include academics, government and private sector statisticians and related experts from

Left to Right: Jim Landwehr (Avaya Labs Research), Karen Kafadar (Indiana U.), Dan Solomon (North Carolina State U.), Board Vice-Chair Roger Hoerl (GE Global Research), Institute), Ruth Williams (U. of California San Diego), Ken Bollen (U. of North Carolina at Chapel Hill), Donna Brogan (Emory U.), Alan Gelfand (Duke U.), Linda Young (U. of Florida), Cliff Spiegelman (Texas A&M U.), Jessica Utts (U. of California Irvine), Phil Kott (RTI International), Board Chair Susan Ellenberg (U. of Pennsylvania), George Williams

Photos from the Annual Board Meeting and the NISS Turkey Fry



Alan Karr, George Williams, Susan Ellenberg and Donna Brogan at the reception and dinner.



Jessica Utts and Mary Ellen Bock at the reception and dinner.



Alan lowering the turkey into the fryer.



Alan carves the turkey.



Bob Rodriguez and Roger Hoerl at the dinner and reception.



Jim Landwehr, Nell Sedransk and Cliff Spiegelman.

Donate to the NISS Fund

Gifts of any size to the NISS Fund enable us to increase our value added to the community, such as helping us pursue new research opportunities, fund a postdoc to attend important workshops and meetings, or for NISS to conduct a new research workshop. Please visit our website at www.niss.org to make your contribution. Your generosity is so greatly appreciated!



A mere 45 minutes later, voilà!



The potluck is ready to begin!



P.O. Box 14006 19 T.W. Alexander Drive Research Triangle Park, NC 27709 919.685.9300 (phone) 919.685.9310 (fax) www.niss.org

NISS/SAMSI Affiliates

Industries

AT&T Labs-Research, Florham Park, NJ Avaya Labs, Basking Ridge, NJ GlaxoSmithKline, Research Triangle Park, NC and Collegeville, PA MetaMetrics, Inc., Durham, NC PNYLAB, LLC, Princeton, NJ RTI International, Research Triangle Park, NC SAS Institute, Cary, NC SPSS, an IBM Company, Chicago, IL Telecordia Technologies, Piscataway, NJ Yahoo! Research Laboratory, Silicon Valley, CA

Government Agencies & National Laboratories

Bureau of Labor Statistics, Washington, DC Energy Information Administration, Washington, DC National Agricultural Statistics Service, Fairfax, VA National Center for Education Statistics, Washington, DC National Center for Health Statistics, Hyattsville, MD National Geospatial-Intelligence Agency, Bethesda, MD National Security Agency, Ft. George W. Meade, MD Office of the Comptroller of the Currency (Treasury Department), Washington, DC US Census Bureau, Washington, DC

Universities

Carnegie Mellon University, Department of Statistics Columbia University, Department of Biostatistics Cornell University, Department of Statistical Science Duke University, Departments of Statistical Science and Mathematics Duke University Medical Center; Department of Biostatistics and Informatics Emory University, Department of **Biostatistics** Florida State University, Department of Statistics Georgia Institute of Technology - School of Mathematics and School of Industrial Systems and Engineering George Mason University, Department of Statistics Georgetown University Medical Center, Department of Biostatistics, Bioinformatics, and Biomathematics Indiana University, Department of Statistics Iowa State University, Department of Statistics Johns Hopkins University, Department of Applied Mathematics and Statistics Medical University of South Carolina, Department of Biostatistics, Bioinformatics and Epidemiology North Carolina State University, Department of Statistics North Carolina State University, Department of Mathematics Oakland University, Department of Mathematics Ohio State University, Department of Statistics

Pennsylvania State University, Department of Statistics Purdue University, Department of Statistics Rice University, Department of Statistics Stanford University, Department of Statistics Texas A&M University, Department of Statistics University of California - Berkeley, Department of Statistics University of Connecticut, Department of **Statistics** University of Florida, Department of **Statistics** University of Georgia, Department of **Statistics** University of Illinois Urbana-Champaign, Department of Statistics University of Maryland Baltimore County, Department of Mathematics & Statistics University of Michigan, Departments of Statistics and Biostatistics University of Missouri-Columbia, Department of Statistics University of North Carolina at Chapel Hill, Department of Statistics and Operations Research University of North Carolina at Chapel Hill, Department of Biostatistics University of North Carolina at Chapel Hill, Department of Mathematics and Statistics University of Pittsburgh - Departments of Statistics and Biostatistics University of South Carolina, Department of **Statistics** Virginia Commonwealth University, Departments of Biostatistics and Statistical Sciences Virginia Polytechnic Institute and State University, Department of Statistics