



WSS NEWS

May 2008

WASHINGTON
STATISTICAL
SOCIETY

Announcement of the 2008 WSS Election

The 2008 WSS election will be held online from **Monday, May 5**, through **Friday, May 30**. Instructions on voting will be sent to members the first week of May. Below are the candidates for this year's election. All WSS members are urged to vote once the balloting begins. The results will be announced at the WSS Annual Dinner on June 25.

Candidates for President (select one)

John L. Eltinge, BLS

John L. Eltinge received a Ph.D. from the Department of Statistics at Iowa State University in 1987. From 1988 to 1999, he served as an assistant and associate professor in the Department of Statistics at Texas A&M University, where he taught statistics and survey sampling, and consulted with the BLS, the NCHS and several Texas state agencies. In 1992-1993, he was an ASA/NSF/BLS Research Fellow. From 1999 to 2004, he worked as the Senior Mathematical Statistician in the Office of Survey Methods Research at the BLS. Since 2004 he has served as the Associate Commissioner for Survey Methods Research at the BLS.

His previous professional service includes work as the program chair for the ASA Survey Research Methods Section (2000); the overall program chair for the Joint Statistical Meetings (2003); Associate Editor for *Journal of the American Statistical Association* (1992-1998); Associate Editor for *The American Statistician* (2000-2004); Associate Editor for *Survey Methodology* (1998-present); and Associate Editor for *Journal of Official Statistics* (2001-present). He is a fellow of the American Statistical Association, and has primary research interests in survey inference, cost structures, optimization, nonresponse and small area estimation.

Marilyn McMillen Seastrom, NCES

Marilyn Seastrom serves as the Chief Statistician and Director of the Statistical Standards Program at the National Center for Education Statistics (NCES) in the US Department of Education. Prior to joining NCES in 1988, she conducted demographic and health analyses for seven years in various components of HHS.

Seastrom received her Ph.D. in Sociology (Demography and Applied Social Statistics) and an MS from the University of Illinois at Urbana-Champaign. In her current position, she led the 2002 revision of the NCES Statistical Standards, which serve as a guide for NCES staff and contractors in conducting surveys and reporting on survey results. She has published both substantive and methodological work using administrative records and sample survey data in government reports and a number of refereed journals.

Seastrom is a Fellow of the American Statistical Association; and has previously served on the Executive Committees of both the Survey Research Methods and Social Statistics Sections of the ASA. She has been a member of the American Statistical Association for 30 years and is also member of the Washington Statistical Society, where she served as Co-Chair Social Arrangements Committee from 1986 to 1987 and as a member of the Herriot Award Committee in 2004. Seastrom has been a member of the OMB Federal Committee on Statistical Methodology since 2003.

Candidates for Methodology Program Chair (select one)

Brian Meekins, BLS

Brian Meekins is currently a research statistician at the Bureau of Labor Statistics. His primary areas of study are latent variable models, measurement error, and telephone survey methodology. Brian received his doctorate from University of Virginia where he worked at the Center for Survey Research for a number of years. Recently, Brian served as the SRMS Section Newsletter Editor from 2004 to 2006.

Y. Micheal Yang, ICF International

Dr. Y. Michael Yang is Chief Statistician at the Survey Research Center at ICF International, a global consulting services firm headquartered in Fairfax, VA. Dr. Yang's research encompasses survey sample design, post-survey data adjustment, and survey data analysis. He has served as senior statistician on many large scale survey projects, responsible for sample design, estimation, and related methodological research. He joined ICF from the National Opinion Research Center (NORC) at the University of Chicago, where he was Senior Statistician in the Statistics and Methodology Department from 2000 to 2007. He has also worked as Senior Methodologist/Statistician at The Gallup Organization.

Candidates for WSS Representative-at-Large (select two)

Jim Knaub, EIA

Jim Knaub has worked for the Federal Government for 30 years, with a variety of science-oriented job titles, and has been a member of ASA and WSS for 18 years. Qualifications include six years on the Mathematics Advisory Committee for the Arlington Public Schools, including one year as the chair, suggesting and following through with projects, and also substantial work as a referee or editor for statistics journals, including the Journal of Official Statistics (JOS), and InterStat (an online journal at <http://interstat.statjournals.net/>), with an emphasis on constructive help for authors. He has presented numerous proceedings papers, most (at least 15) for JSM sessions, and has twice chaired at the JSM, presented two WSS Seminars, and has also published articles on InterStat. In June 2006 his book review of *Estimation in Surveys with Nonresponse*, by Carl-Erik Sarndal and Sixten Lundstrom (Wiley), appeared in JOS (<http://www.jos.nu/Articles/abstract.asp?article=222351>). Jim has written two entries for the Sage *Encyclopedia of Measurement and Statistics*, and two other entries for the yet-to-be-released Sage *Encyclopedia of Survey Research Methods*, including one entry on cutoff sampling. Further, Jim has presented several times to the ASA Energy Committee; is a former member of the ASA Human Rights Committee; and has been a WSS science fair judge and presented for WSS to scouts. At the Electric Power Division within the Energy Information Administration, Jim developed estimation procedures and samples for electric power surveys with emphasis on integrating with the total survey systems, including data quality and error measurement considerations. Over a number of years he has developed an extensive system, and as lead statistician in that area, now organizes the work of others. His first Govt. supervisor called Jim a problem solver and self-starter, and he continues to strive for that. Jim considers himself well rounded due to other interests. (Come hear the band: <http://www.viennacommunityband.org/concertschedule.php>! Jim plays trombone in a concert band.)

Tim Kennel, U.S. Census Bureau

Tim Kennel enjoys learning about statistics, using statistics to improve surveys, and interacting with the wonderful Washington statistical community. Presently, Tim is a second year doctoral student studying survey methodology at the University of Maryland and a Mathematical Statistician at the US Census Bureau. When he is not working or studying, Tim likes cooking, going to museums, and spending time with his delightful wife. For the past four years, Tim has humbly served on the WSS membership committee with joy. He hopes to serve this community in a new way.

Elizabeth H. Margosches, US EPA

Elizabeth Margosches would like the opportunity to be more active with the WSS and serving as a representative at large seems to her a good way to start and a way her other experiences can best inform her service. Currently Statistician in the Existing Chemicals Assessment Branch, Risk Assessment Division, Office of Pollution Prevention and Toxics (OPPT), Office of Prevention, Pesticides and Toxic Substances (OPPTS), US Environmental Protection Agency (USEPA), Washington, DC, Dr. Margosches has been in the Agency since 1980 in various risk-assessment-related roles and serving on a variety of Agency and external workgroups and committees. Coming to the EPA fresh from her PhD and MPH in Biostatistics at the University of Michigan, she joined the Washington Statistical Society soon after (ASA membership was not a requirement at the time). Then professionally junior, she was encouraged by the WSS's offshoot of a group (the Washington Alliance of Statisticians) to mentor statisticians in the early years of their careers. Eventually, she did join the American Statistical Association (where she has served on the Committee on Women in Statistics since 2003, as chair in 2005) and the Government Statistics Section (this latter about 5 years ago). She also belongs to the International Biometric Society/Eastern North American Region (ENAR; Regional Advisory Board, 2002-2004), the Caucus for Women in Statistics (Representative at Large, 1992-1994; President, 1998 with 1997-2000 term), and the Association for Women in Science/DC Chapter (Treasurer 1988-2003).

Darryl Creel, RTI International

Darryl V. Creel is a Research Statistician at RTI International and has worked there for three and a half years. Prior to working at RTI, he worked at Mathematica Policy Research. In 1997, he received an M.S. in Statistical Science from George Mason University and became a member of the Washington Statistical Society. His general research interest is the survey process. His specific research areas of interest are imputation, nonresponse adjustments, and the analysis of survey data. For a decade, he has enjoyed the benefits of being a WSS member and would like the opportunity to give something back to the statistical community.

Candidate for Treasurer (select one)

Jane Li, Westat

Jane Li is a sampling statistician at Westat since August 2007. She received her master's degree in Economics from University of Delaware in December 2002. After that she was admitted to the doctoral program in the Joint Program in Survey Methodology at University of Maryland and received her Ph.D. in December 2007.

WSS and Other Seminars

(All events are open to any interested persons)

May

- 2 Fri. **Statistical issues in disease surveillance: A case study from ESSENCE**

- 5 Mon. **Statistical Issues Arising in the Interpretation of a Measure of Relative Disparity Used in Educational Funding: The Zuni School District 89 Case**

- 13 Tues. **Multivariate Event Detection and Characterization**

- 15 Thur. **President's Invited Seminar: What's Up at the ASA?**

- 16 Fri. **Bayesian Dose-finding Trial Designs for Drug Combinations**

June

- 10 Tues. **Nonresponse Adjustments in Survey Applications**

- 11 Wed. **Save the Date: *Elections and Exit Polls***

- 17 Tues. **Recent Developments in Address-based Sampling**

- 26 Thurs. **Multiple Frame Surveys: Lessons from CBECS Experience**

Also available on the Web at the following URL: <http://www.scs.gmu.edu/~wss/>

Annual Dinner

This year's WSS Annual Dinner is at the MeiWah restaurant in Chevy Chase, Maryland (Friendship Heights Metro), Wednesday, June 25, 2008. The Gertrude Cox Award winner is Dr. Thomas Lumley from the University of Washington. Dr Lumley will speak at the dinner. The title of the talk is "Open Source Statistical Software: Why? When? Where?". The price for the dinner is \$45 person. The registration flyer is attached.

Announcement

SIGSTAT Topics for Spring 2008

May 21, 2008: Survival Models in SAS: PROC PHREG – Part 2

Continuing the series of talks based on the book “Survival Analysis Using the SAS System: A Practical Guide” by Paul Allison begun in October 2007, we’ll continue with Chapter 5: Estimating Cox Regression Models with PROC PHREG.

Topics covered are: Tied data

June 18, 2008: Survival Models in SAS: PROC PHREG – Part 3

Continuing the series of talks based on the book “Survival Analysis Using the SAS System: A Practical Guide” by Paul Allison begun in October 2007, we’ll continue with Chapter 5: Estimating Cox Regression Models with PROC PHREG.

Topics covered are: Time-Dependent Covariates

SIGSTAT is the Special Interest Group in Statistics for the **CPCUG**, the Capital PC User Group, and **WINFORMS**, the Washington Institute for Operations Research Service and Management Science.

All meetings are in Room S3031, 1800 M St, NW from **12:00 to 1:00**. Enter the South Tower & take the elevator to the 3rd floor to check in at the guard’s desk. First-time attendees should contact Charlie Hallahan, 202-694-5051, hallahan@ers.usda.gov, and leave their name. Directions to the building & many links of statistical interest can be found at the **SIGSTAT** website, <http://www.cpcug.org/user/sigstat/>.

Announcement

Georgetown University Department of Biostatistics, Bioinformatics and Biomathematics

Are you interested in genetics, bioterrorism, international health, bioinformatics, epidemiology or health policy? Get the analytic tools to meet the demands of the 21st century!

You are invited to apply for the Master of Science Program in Biostatistics with tracks in Bioinformatics and/or Epidemiology for Fall 2008. Application deadline is June 30, 2008.

Graduates with an MS degree in Biostatistics go on to successful and lucrative positions in academic centers, pharmaceutical companies, biotech companies and private consulting firms.

For more information, visit: <http://www1.georgetown.edu/gumc/departments/biostatistics/> or e-mail Caroline at ctw26@georgetown.edu.

Apply now for Fall 2008 at: http://grad.georgetown.edu/pages/apply_online.cfm

Announcement

106th Meeting of the Committee on National Statistics

The National Academies, NAS building, 2100 C St., NW, Washington, DC

Friday, May 9, 2008 Open Session - Public Seminar and Reception

2:00 Light refreshments for seminar guests (NAS Great Hall)

Seminar (Auditorium)

2:30 Welcome and Introduction-Bill Eddy, CNSTAT and Carnegie Mellon University

2:35 Developments at the OMB Statistical and Science Policy Office-Katherine Wallman, Chief Statistician

2:45 Featured Topic: 25 Years (and Counting) of Cognitive Survey Research: Accomplishments, Current Work, and Future Opportunities

Panel:

* Roger Tourangeau, Joint Program in Survey Methodology, University of Maryland

* Gordon Willis, Applied Research Program, National Cancer Institute

* Fred Conrad, Institute for Social Research, University of Michigan

3:45 Floor discussion

4:00 Reception (Great Hall)

5:00 Adjourn

Abstract: The May 2008 seminar celebrates the 25th anniversary of the NSF-funded "advanced research seminar" organized under CNSTAT. This intense activity brought together survey researchers and cognitive psychologists to brainstorm about how cognitive research insights could enhance the design of survey questionnaires to improve the completeness and accuracy of reporting. The seminar's report, *Cognitive Aspects of Survey Methodology* (NRC, 1984), boosted efforts by statistical agencies and other organizations to adopt cognitive questionnaire testing as standard best practice for survey work.

A distinguished panel will offer their observations on cognitive survey research—past, present, and future. Roger Tourangeau, director of the Joint Program in Survey Methodology, University of Maryland at College Park, helped produce the 1984 CASM report; he will review the history and past accomplishments of the field. Gordon Willis, who used cognitive methods extensively at the National Center for Health Statistics and is now Cognitive Psychologist at the National Cancer Institute, will discuss current cognitive survey work. Fred Conrad, research associate professor, Institute for Social Research, University of Michigan, will challenge the field to strengthen its grounding in evidence-based research as it looks to the future.

Special guests will include the chair and members of the original CASM advanced research seminar: Judy Tanur, SUNY Stony Brook (chair), Tom Jabine, CNSTAT consultant, Norman Bradburn, NORC, Monroe Sirken, NCHS retired, and Miron Straf, National Academies.

Note: All venues are handicapped-accessible. There is limited first-come, first-served parking in the visitors' lot on 21st St. The nearest Metro station is Foggy Bottom at 23rd and I Sts., NW (Blue and Orange Lines).

Please RSVP by May 6 to Bridget Edmonds at 202-334-3096 or cnstat@nas.edu <<mailto:cnstat@nas.edu>> .

Program Announcement

- Title:** **Statistical issues in disease surveillance: A case study from ESSENCE**
- Speaker:** Cara Olsen, PhD, Biostatistics Consulting Center (CIV, USUHS)
- Date/Time:** Friday, May 2, 2008 / 10:00 -11:00 a.m.
- Location:** Georgetown University Medical Center, Building D, 4000 Reservoir Rd., NW, Warwick Evans Conference Room, Washington, DC 20007
- Sponsor:** Department of Biostatistics, Bioinformatics and Biomathematics
- Abstract:** Syndromic surveillance systems attempt to monitor the burden of disease in communities in real time, using health-related data and tools from statistics, epidemiology, informatics, and other disciplines. A potential benefit of such surveillance is early detection and tracking of infectious disease outbreaks.

The Electronic Surveillance System for the Early Notification of Community-based Epidemics (ESSENCE) is a syndromic surveillance system that monitors outpatient visits to military medical treatment facilities. This study examines whether ESSENCE can detect more infectious disease outbreaks, and detect them earlier, using joint monitoring of laboratory test orders and outpatient visit data rather than outpatient visit data alone. Statistical issues that arise from this question include which aberration detection algorithm is best suited to these data sources, how to quantify the tradeoffs among sensitivity, specificity and timeliness for detecting outbreaks, and how to monitor information from multiple data sources simultaneously.

For information, please contact Caroline Wu at 202-687-4114 or ctw26@georgetown.edu

Program Announcement

Title: **Statistical Issues Arising in the Interpretation of a Measure of Relative Disparity Used in Educational Funding: The Zuni School District 89 Case**

Speaker: Joseph L. Gastwirth, Department of Statistics, George Washington University

Discussant: Marc Rosenblum, Office of the General Counsel and Chief Economist, Equal Employment Opportunity Commission

Chair: Michael L. Cohen, Committee on National Statistics

Date/Time: Monday, May 5, 2008 / 12:30 to 2:00 p.m.

Location: Bureau of Labor Statistics Conference Center. To be placed on the seminar attendance list at the Bureau of Labor Statistics, you need to email your name, affiliation and seminar name to wss_seminar@bls.gov (note that there is an underscore after 'wss') by noon at least two days in advance of the seminar or call 202-691-7524 and leave a message with this information. Bring a photo ID to the seminar. BLS is located at 2 Massachusetts Ave., NE. Take the Red Line to Union Station.

Sponsor: WSS Section on Public Policy

Abstract: This seminar will discuss statistical issues that arose in recent cases. The first case concerns the interpretation of a formula Congress wrote when it revised a law that provides funds for educating children in areas with a large federal presence (e.g. major research lab). Because federal land is not subject to local real estate tax, the primary source of funding education, the law is intended to assist the relevant school districts. We will discuss the statute and the various interpretations that arose during the proceedings and the justifications provided. A counter-example to one of the assertions made by the lawyers at the Supreme Court hearing, which appears to have been accepted by the Court's majority, will also be presented.

Program Announcement

- Title:** **Multivariate Event Detection and Characterization**
- Speaker:** Daniel B. Neill, Carnegie-Mellon University
- Chair:** Myron Katzoff, CDC/National Center for Health Statistics
- Date/Time:** Tuesday, May 13, 2008 / 12:30 to 2:00 p.m.
- Location:** Bureau of Labor Statistics Conference Center. To be placed on the seminar attendance list at the Bureau of Labor Statistics, you need to email your name, affiliation and seminar name to wss_seminar@bls.gov (note that there is an underscore after 'wss') by noon at least two days in advance of the seminar or call 202-691-7524 and leave a message with this information. Bring a photo ID to the seminar. BLS is located at 2 Massachusetts Ave., NE. Take the Red Line to Union Station.
- Sponsor:** Defense and National Security Section
- Abstract:** We present the multivariate Bayesian scan statistic (MBSS), a general framework for event detection and characterization in multivariate spatial time series data. MBSS integrates prior information and observations from multiple data streams in a principled Bayesian framework, computing the posterior probability of each type of event in each space-time region. MBSS learns a multivariate Gamma-Poisson model from historical data, and models the effects of each event type on each stream using expert knowledge or labeled training examples. We evaluated MBSS on various disease surveillance tasks, detecting and characterizing disease outbreaks injected into three streams of Pennsylvania medication sales data. We demonstrated that MBSS can be used both as a "general" event detector, with high detection power across a variety of event types, and a "specific" detector that incorporates prior knowledge of an event's effects to achieve much higher detection power. MBSS has many other advantages over previous event detection approaches, including efficient computation and easy interpretation and visualization of results, and allows faster and more accurate detection by integrating information from the multiple streams. Most importantly, MBSS can model and differentiate between multiple event types, thus distinguishing between events requiring urgent responses and other, less relevant patterns in the data. This talk will present an overview of the MBSS framework, and compare MBSS to other recently proposed multivariate detection approaches. Time permitting, I will also discuss how incremental learning (both passive and active) can be incorporated into the MBSS framework and used to improve detection performance, and consider extensions of MBSS to more general pattern detection problems.

Program Announcement

- Title:** President's Invited Seminar: What's Up at the ASA?
- Chair:** Michael P. Cohen, WSS President
- Speaker:** Ron Wasserstein, ASA Executive Director
- Date/Time:** Thursday, May 15, 2008 / 12:30 pm to 2:30 pm
- Location:** Bureau of Labor Statistics, Conference Center. To be placed on the seminar list attendance list at the Bureau of Labor Statistics you need to e-mail your name, affiliation, and seminar name to wss_seminar@bls.gov (underscore after 'wss') by noon at least 2 days in advance of the seminar or call 202-691-7524 and leave a message. Bring a photo ID to the seminar. BLS is located at 2 Massachusetts Avenue, NE. Use the Red Line to Union Station.
- Sponsor:** WSS
- Abstract:** ASA Executive Director Ron Wasserstein will provide a brief update on activities and directions of the association. However, most of the session will be devoted to questions and comments from the participants. Among the many things we could discuss:
- What needs to be done to attract and retain statisticians in government service, and what role could the ASA play?
 - What benefits and services could the ASA provide that would increase its attractiveness to public and private sector members in this area?
 - Where do you see the profession heading in the next few years, and what should the ASA be doing?
- Ron has his opinions, of course, but is most interested in hearing yours.

Program Announcement

- Title:** Bayesian Dose-finding Trial Designs for Drug Combinations
- Speaker:** Guosheng Yin, Ph.D., Assistant Professor, Department of Biostatistics, M. D. Anderson Cancer Center
- Date/Time:** Friday, May 16, 2008 / 10:00 -11:00 a.m.
- Location:** Georgetown University Medical Center, Lombardi Comprehensive Cancer Center, 3900 Reservoir Rd., NW, Research Building, Conference Room E501, Washington, DC 20007
- Sponsor:** Department of Biostatistics, Bioinformatics and Biomathematics
- Abstract:** Treating patients with a combination of agents is becoming commonplace in cancer clinical trials, with biochemical synergism often the primary focus. In a typical drug combination trial, the toxicity profile of each individual drug has already been thoroughly studied in the single-agent trials, which naturally offers rich prior information. We propose Bayesian adaptive designs to search for the maximum tolerated dose combination. We continuously update the posterior estimates for the toxicity probabilities of the combined doses. By reordering the dose toxicities in the two-dimensional probability space, we adaptively assign each new cohort of patients to the most appropriate dose. Dose escalation, de-escalation or staying the same is determined by comparing the posterior estimates of the toxicity probabilities of combined doses and the prespecified toxicity target. We conduct extensive simulation studies to examine the operating characteristics of the design and illustrate the proposed method under various practical scenarios.

For information, please contact Caroline Wu at 202-687-4114 or ctw26@georgetown.edu

Program Announcement

- Title:** Nonresponse Adjustments in Survey Applications
- Chair:** Nancy Bates, U.S. Census Bureau
- Speakers:** Frauke Kreuter, Joint Program in Survey Methodology, University of Maryland
Trena Ezzati-Rice, Agency for Healthcare Research and Quality
- Discussant:** Keith Rust, Westat and JPSM
- Date/Time:** Tuesday, June 10, 2008 / 12:30 - 2:00 p.m.
- Location:** Bureau of Labor Statistics, Conference Center. To be placed on the seminar attendance list at the Bureau of Labor Statistics you need to e-mail your name, affiliation, and seminar name to wss_seminar@bls.gov (underscore after 'wss') by noon at least two days in advance of the seminar, or call 202-691-7524 and leave a message. Bring a photo ID to the seminar. BLS is located at 2 Massachusetts Avenue, NE. Use the Red Line to Union Station.
- Sponsor:** WSS Methodology Program

Abstract (Kreuter):

Using Proxy Measures and Other Correlates of Survey Outcomes to Adjust for Nonresponse: Examples from Multiple Surveys Nonresponse weighting is a commonly used method to adjust for bias due to unit nonresponse in surveys. Theory and simulations show that, in order to effectively reduce bias without increasing variance, a covariate used for nonresponse weighting adjustment needs to be highly associated with both response and the survey outcome. In practice, these requirements pose a challenge that is often overlooked. Recently some surveys have begun collecting supplementary data, such as interviewer observations and other proxy measures of key survey outcomes. These variables are promising candidates for nonresponse adjustment because they should be highly correlated with the actual outcomes. In the present study, we examine the extent to which traditional covariates and new proxy measures satisfy the weighting requirements for the National Survey of Family Growth, the Medical Expenditure Survey, the U.S. National Election Survey, the European Social Surveys and the University of Michigan Transportation Research Institute Survey. We provide empirical estimates of the association between proxy measures and the likelihood of response as well as the actual survey responses. We also compare unweighted and weighted estimates under various nonresponse models. Results show the difficulty of finding suitable covariates and the need to improve the quality of proxy measures.

Abstract (Ezzati-Rice):

Assessment of the Impact of Health Variables on Nonresponse Adjustment in the Medical Expenditure Panel Survey The Medical Expenditure Panel Survey (MEPS) is a large complex sample survey, designed to provide nationally representative annual estimates of health care use, expenditures, sources of payment, and insurance coverage for the U.S. civilian non-institutionalized population. A new panel of households is selected each year for the MEPS from households that responded to the previous year's National Health Interview Survey (NHIS). Nonresponse is a common problem in household sample surveys. To compensate for nonresponse and to reduce the potential bias of the survey estimates, two separate nonresponse adjustments are performed in development of analytic weights in MEPS. The first, the focus of this presentation, is an adjustment for dwelling unit (DU) level nonresponse to account for nonresponse among those households subsampled from NHIS for the MEPS. The adjustment is carried out using socio-economic, demographic, and health variables that are available for both respondents and nonrespondents. In this study, we examine the impact of health variables on the MEPS DU level nonresponse weight adjustment. Response propensity scores are calculated based on logistic regression models and quintiles of the propensity scores are used to adjust the MEPS base weights. Comparisons of the nonresponse adjusted weights and selected survey variables with and without inclusion of health variables as a nonresponse adjustment covariate are discussed.

Program Announcement

Save the Date

Wednesday, June 11th 12:00 noon-2:00pm

Book Signing at Reiter's Books

Wendy Alvey and Fritz Scheuren
Elections and Exit Polls

Lite refreshments (wine, cheese, and soft drinks) provided.

Both authors will be there:

Reiter's Books
1990 K Street NW
Washington DC

Sponsor: WSS Human Rights Section and WSS Public Policy Section

Look forward to seeing you there

Program Announcement

- Title:** Recent Developments in Address-based Sampling
- Chair:** Meena Khare, NCHS
- Speakers:** Mansour Fahimi, Marketing Systems Group
- Discussant:** Sylvia Dohrmann, Westat
- Date/Time:** Tuesday, June 17, 2008 / 12:30 - 2:00 p.m.
- Sponsor:** WSS Methodology Program
- Location:** Bureau of Labor Statistics, Conference Center Room 8
- Attendance:** In order to be placed on the seminar list attendance list at the Bureau of Labor Statistics you need to e-mail your name, affiliation, and seminar name to wss_seminar@bls.gov (underscore after 'wss') by noon at least 2 days in advance of the seminar or call 202-691-7524 and leave a message. Bring a photo ID to the seminar. BLS is located at 2 Massachusetts Avenue, NE. Use the Red Line to Union Station.
- Abstract:** Increasingly, survey researchers are reverting back to address-based methodologies to reach the general public for survey administration and related commercial applications. Essentially, there are three main factors for this change: evolving coverage problems associated with telephone-based methods; eroding rates of response to telephone contacts; and on the other hand, recent improvements in the databases of household addresses available to researchers. This presentation provides an assessment of these three factors along with an overview of the structure of the Delivery Sequence File (DSF) of the USPS that is often used for construction of address-based sampling frames. Moreover, key enhancements available for the DSF will be discussed. While reducing undercoverage bias particularly in rural areas where more households rely on P.O. Boxes and inconsistent address formats such enhancements enable researcher to develop more efficient sample designs as well as broaden their analytical possibilities through an expanded set of covariates for hypothesis testing and statistical modeling tasks.

Program Announcement

- Title:** Multiple Frame Surveys: Lessons from CBECS Experience
- Speaker:** Stephanie Eckman and Katie Dekker, NORC
- Discussant:** Mike Brick, Westat and JPSM
- Chair:** Eugene Burns, Energy Information Administration
- Date/Time:** Thursday, June 26, 2008 / 12:30 - 2:00 p.m.
- Location:** Bureau of Labor Statistics Conference Center. To be placed on the seminar attendance list at the Bureau of Labor Statistics, you need to email your name, affiliation and seminar name to wss_seminar@bls.gov (note that there is an underscore after 'wss') by noon at least two days in advance of the seminar or call 202-691-7524 and leave a message with this information. Bring a photo ID to the seminar. BLS is located at 2 Massachusetts Ave., NE. Take the Red Line to Union Station.
- Sponsor:** WSS Data Collection Methods Section
- Abstract:** Because there is no single frame of buildings in the country, the Commercial Building Energy Consumption Survey relies on federal databases, commercially-available databases and field listing to construct a sampling frame. The 2007 round of CBECS, which NORC is currently fielding on behalf of the Energy Information Administration of the DOE, sampled from seven frames. The difficulty in sampling simultaneously from many frames is in adjusting for the overlap among them. To calculate the correct weights, we must know the true probability of selection for all selected cases: for each selected building we must either know its probability of selection from each of the frames to which it belongs or we must remove it from all of the frames but one. We will present the steps we took to remove these overlaps before the sample was fielded. Despite our best efforts, duplicates were discovered in the field: we will discuss the ways in which we modified the probabilities of selection once data collection was underway. We must also confront the fact that undetected duplicates remain, and that some of the probabilities of selection will not be correct. The CBECS experience can help other surveys to decide whether incorporating additional frames is worth the difficulty. Our findings also apply to dual-frame phone surveys, where it may not be possible to fully deduplicate the frames.

JPSM SHORT COURSES

JPSM HOME PAGE: <http://www.jpsm.org> Click on "Short Courses"

INFORMATION, REGISTRATION, PAYMENT AND CANCELLATION:

<https://projects.isr.umich.edu/jpsm/>

SPONSOR AFFILIATES: <https://projects.isr.umich.edu/jpsm/sponsorlist.cfm>

May 5-6, 2008 (Open for registration)

Sampling and Estimation for Establishment Surveys

Richard L. Valliant and Phill Kott

Registration Deadline: April 25, 2008

Information: <https://projects.isr.umich.edu/jpsm/materials/2008-0506.html>

May 20-21, 2008 (Open for registration)

Applied Structural Equation Modeling

Patrick Sturgis

Registration Deadline: May 6, 2008

Information: <https://projects.isr.umich.edu/jpsm/materials/2008-0520.html>

May 27-28, 2008 (Not yet open for registration) Focus Groups from Start to Finish David Morgan

Primary Funding for JPSM is from the Interagency Council on Statistical Policy

Students' Corner

Jill Montaquila, former president of the Washington Statistical Society, recently forwarded the following information to me. The **Consortium for the Advancement of Undergraduate Statistics Education (CAUSE)** has announced its second biennial **Undergraduate Statistics Project Competition (USPROC)**. The competition is open to undergraduate students globally. Under the guidance of at least one instructor sponsor, working in groups of up to three people, students participating in the competition will present posters showcasing a study in which they have applied statistical methods to real-world data. The top four projects will earn monetary awards of \$750 for first place, \$500 for second place, \$250 for third place and \$100 for fourth place.

Students conducting their projects in Spring of 2008 are encouraged to submit their projects as early as May 2008. CAUSE will be accepting submissions for USPROC starting May 5th, 2008, up to January 26th, 2009. Winners will be announced March 7th, 2009. To learn more about the competition, including project scope, the criteria by which projects will be judged, and guidelines for submission, see this document:

http://bist.pbwiki.com/f/Guidelines_USPROC_2009.doc

Or visit the CAUSE website: <http://www.causeweb.org/usproc/>

Thanks for this information, Jill!

If you're a student attending a high school or middle school in the Washington, D.C. area, here's an announcement for a competition for you to consider: the **Curtis Jacobs Memorial Prize For Outstanding Statistics Project**. Students must have a teacher or an advisor who will cover material on statistics during the 2007/2008 academic year, and must submit a five- to ten-page typewritten report of their statistical project. The deadline is May 18, 2008.

See this web page for details: <http://www.scs.gmu.edu/~wss/jacobs.html>

If you don't have plans for the summer, consider taking a summer course in statistics. I recently received in the mail an announcement for the **ICPSR Summer Program in Quantitative Methods of Social Research**. It has a fairly large and varied course offering, ranging from undergraduate to graduate-level and beyond. Most of them are held in Ann Arbor, Michigan, at the University of Michigan. For more details, see this web page: <http://www.icpsr.umich.edu/summerprog>. I note that by the time you read this, the deadline will unfortunately have passed for most of the offerings. But perhaps you can still take one of the 3- or 5-day workshops.

If you're in the situation where there's a topic that you simply must learn about, then perhaps you might consider taking one of the workshops. For example, maybe you have sunk a lot of time into your thesis when you realize that there is some statistical topic that you need to learn about, but that isn't offered by your department. Perhaps you can take one of the summer workshops, then.

Registration for the Joint Statistical Meetings opens May 1, 2008. See this web page for details:

<http://www.amstat.org/meetings/jsm/2008/>

This month, I'll present the second of this two-part tutorial on getting started with C++. Last month, we installed **Dev-C++**, a full-featured C++ integrated development environment. Then we used Dev-C++ to compile a small "Hello World" program, and to compile **Newmat**, a C++ implementation of useful linear algebra classes and functions into a statically linked library. This month, we'll make a small program that demonstrates some of the possibilities with Newmat and with C++ in general. Afterwards, I'll give some suggestions for further projects. I will assume that you have already gotten through the first part of this tutorial. Let's continue, then.

6. Create New C++ Project: Calling a Function, and Matrix Manipulations

Here, we'll compile a simple C++ program that demonstrates how to use the Newmat source code to perform linear algebra computations, using Dev-C++. This program first calls a function that reads data from a plain text file into a Newmat matrix. It then uses matrix transpose and inverse operations to compute a least-squares estimate, and then finally it demonstrates how to compute Singular Value Decomposition using Newmat.

1. Repeat Steps III(B) – III(C) from Part I of this tutorial to create a new C++ **Console Program** project named **matrixdemo**. Save the project into the `matrixdemo` folder we created in step II, accepting the default name of `matrixdemo.dev`.
2. As in Step III(D), a small default program appears. Modify it so that it looks like this (if you're reading an electronic copy of this tutorial, copying from this page and pasting into Dev-C++ would be most convenient here):


```
#include "matrixdemo.h"
int main(int argc, char *argv[])
{
    // Read data matrix into Matrix 'rawdata'
    Matrix rawdata;
    ReadTextFileIntoMatrix(argv[1],rawdata);

    // Copy the first column of 'rawdata' into ColumnVector 'Y'
    // Copy the rest of 'rawdata' into Matrix 'X'
    ColumnVector Y = rawdata.SubMatrix(1,10,1,1);
    Matrix X = rawdata.SubMatrix(1,10,2,4);

    // Compute Bhat = inv(X'*X)*X*Y.
    // If 'X' is a Newmat matrix, then
    // X.t() is the transpose and X.i() is the inverse
    // (X.t()*X).i() is therefore the inverse of (X'*X)
    Matrix Bhat = (X.t()*X).i() * X.t() * Y;
    cout << "Least Squares Estimates:" << endl;
    cout << "Y = " << endl << Y << endl;
    cout << "X = " << endl << X << endl;
    cout << "Bhat = " << endl << Bhat << endl;

    // Compute Singular Value Decomposition of X
    DiagonalMatrix D;
    Matrix U,V;
    SVD(X,D,U,V);
    cout << "Singular Value Decomposition of X:" << endl;
    cout << "U = " << endl << U << endl;
    cout << "D = " << endl << D << endl;
    cout << "V = " << endl << V << endl;

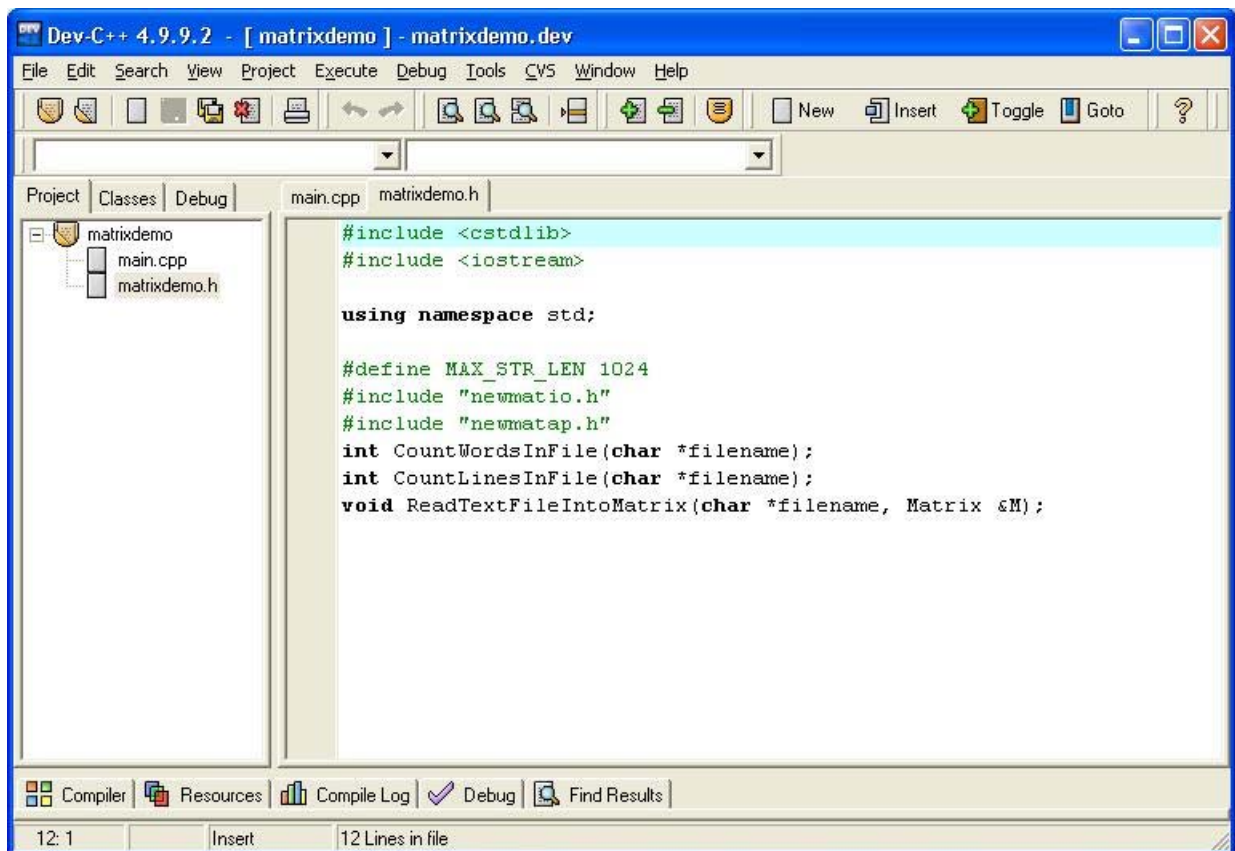
    system("PAUSE");
    return EXIT_SUCCESS;
}
```

3. As in Step III(E), save the program as a file named `main.cpp` in the `matrixdemo` folder.
4. Create a new source code file by selecting **Project** → **New file**. This opens up a new tabbed pane with a temporary name like `[*] Untitled1`. Cut and paste the lines of code shown below into this new tabbed pane, and then click on the **Save** button  to save the content into a file. Save the file in the `matrixdemo` folder, naming it `matrixdemo.h`. This *header file* contains two lines that include two header files from the Newmat source code, which are required if we want to use Newmat. The last three lines are *function declarations*.

```
#include <cstdlib>
#include <iostream>

using namespace std;

#define MAX_STR_LEN 1024
#include "newmat.h"
#include "newmatap.h"
int CountWordsInFile(char *filename);
int CountLinesInFile(char *filename);
void ReadTextFileIntoMatrix(char *filename, Matrix &M);
```



When you're done, the Dev-C++ window should look like the above picture. Note that there are now two tabs above the editor pane, one labeled `main.cpp` and one labeled `matrixdemo.h`.

Clicking on either tab allows you to edit the corresponding file.

- Repeat Step VI(D), saving the following lines of code into a new file named `CountWordsInFile.cpp`. This function's name tells you its function: it counts the words in a file. `MAX_STR_LEN` is a *macro* which was defined in `matrixdemo.h`. And the `fscanf` function reads data from a formatted line into a variable.

```
#include "matrixdemo.h"
int CountWordsInFile(char *filename) {

    int count = 0;
    char dummy[MAX_STR_LEN];
    FILE *fp = fopen(filename, "r");
    while (fscanf(fp, "%s", dummy) != EOF)
        count++;
    fclose(fp);
    return count;
}
```

- Repeat Step VI(D), saving the following lines of code into a new file named `CountLinesInFile.cpp`. This function is very similar to the preceding function, but it counts the lines rather than the words in a file. Compare it to the preceding function; the `fgets` function reads in a line with a specified maximum length.

```
#include "matrixdemo.h"
int CountLinesInFile(char *filename) {

    int count = 0;
    char dummy[MAX_STR_LEN];
    FILE *fp = fopen(filename, "r");
    while (fgets(dummy, MAX_STR_LEN, fp) != NULL)
        count++;
    fclose(fp);
    return count;
}
```

- Repeat Step VI(D), saving the following lines of code into a new file named `ReadTextFileIntoMatrix.cpp`. The “Matrix &M” in the argument list of this function, specifically the ampersand, indicates that matrix M is being passed *by reference*. This is an efficient way to pass relatively large objects (such as a matrix) between functions. The `atof` function converts a character string to a floating point numeric value.

```
#include "matrixdemo.h"
void ReadTextFileIntoMatrix(char *filename, Matrix &M) {

    // Count number of words and rows in text file.
    int words    = CountWordsInFile(filename);
    int rows     = CountLinesInFile(filename);
    int columns  = words/rows;

    // Read values from text file into Matrix
    M.ReSize(rows, columns);
    char dummy[MAX_STR_LEN];
    FILE *fp = fopen(filename, "r");
    for (int i=1; i<= rows; i++) {
```



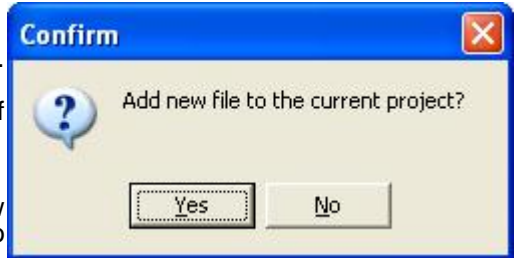
```

        for (int j=1; j<=columns; j++) {
            fscanf(fp,"%s",dummy);
            M(i,j) = (Real) atof(dummy);
        }
    }
    fclose(fp);
}
}
8.

```

OK, this one is different. Select **File** → **New** → **Source File**.

A dialog box labeled Confirm will appear, asking whether to add the new file to the current project. Click on the **No** button. A new tabbed editor pane will appear, with a label like "Untitled1." Since we have chosen not to make this file part of the **matrixdemo** project, an entry for it will not appear in the Project pane on the left.



Copy the following lines of data into the new "Untitled" tabbed editor pane, and then save it into new file in the **matrixdemo** folder named **data.txt**.

This is sample data that our program will read in. The first column will be the observed data vector, **Y**, while the 2nd through 4th columns will be the design matrix, **X**.

```

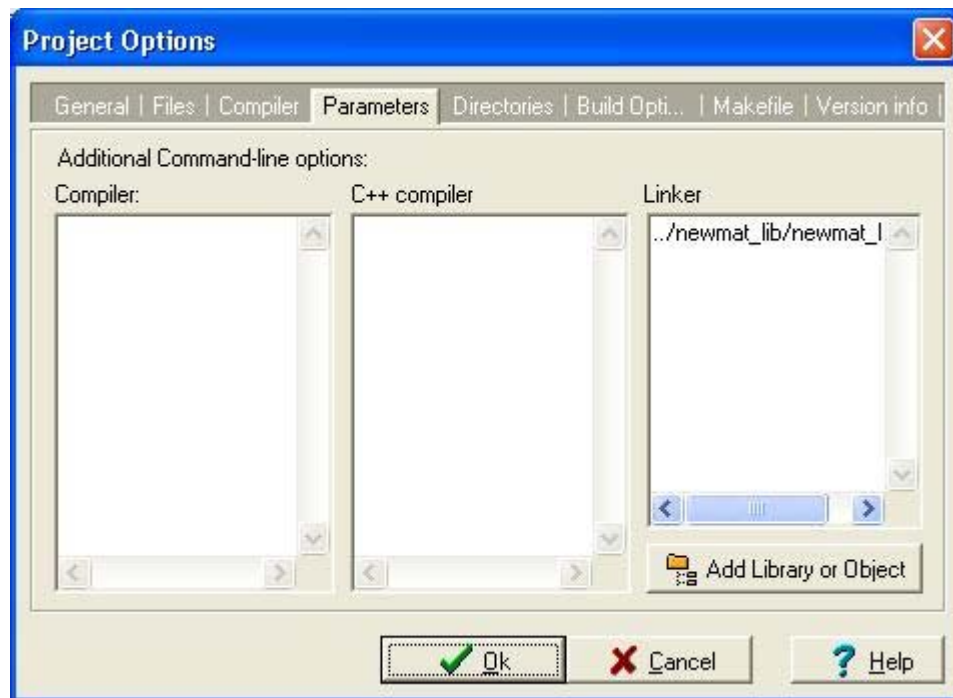
-9.1853    1.0000   -9.0000   -2.0000
-5.0942    1.0000   -7.0000   -2.0000
 3.1270    1.0000   -5.0000    3.0000
 7.9134    1.0000   -3.0000    3.0000
 6.6324    1.0000   -1.0000   -2.0000
10.0975    1.0000    1.0000   -2.0000
19.2785    1.0000    3.0000    3.0000
23.5469    1.0000    5.0000    3.0000
22.9575    1.0000    7.0000   -2.0000
26.9649    1.0000    9.0000   -2.0000

```


Note that you really do *not* want to make this file part of the project. I tried that, and the compiler attempted to compile this text file as if it were C++ source code, generating an error!

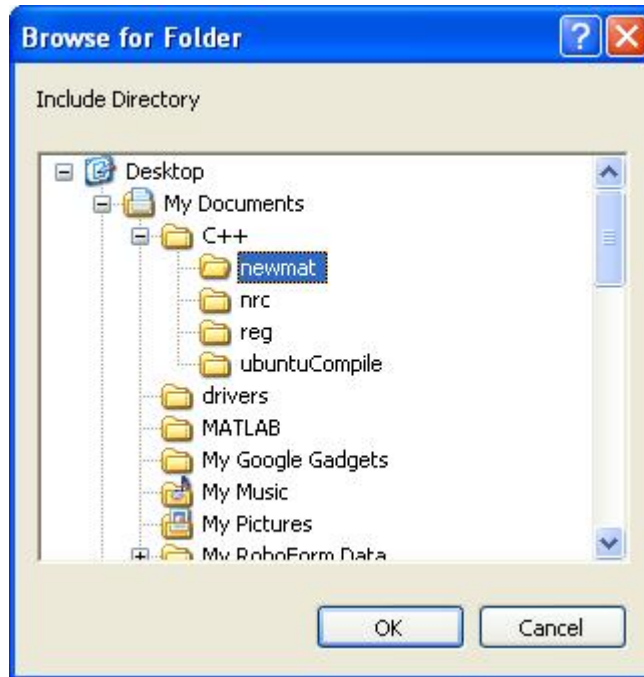
9. We need to tell the compiler where to find the Newmat static library, as well as the Newmat header files. Here's how to do this.

1. Select **Project** → **Project Options**. This will pop up the **Project Options** window.
2. In the **Project Options** window, click on the tab labeled **Parameters** and do the following.
 1. Click on the button labeled **Add Library or Object**, in the lower right corner. This pops up a file browser window labeled **Open**.
 2. In the **Open** window, browse to the **newmat** folder that we created in Step II. Select the file named **newmat.a**, that we created in step V(E). Then click on the **OK** button. This dismisses the **Open** window, and causes an entry for the Newmat library to be entered into the list under **Linker**, in the **Project Options** window. This tells the compiler where to find the Newmat library. The **Project Options** window should now look like this:

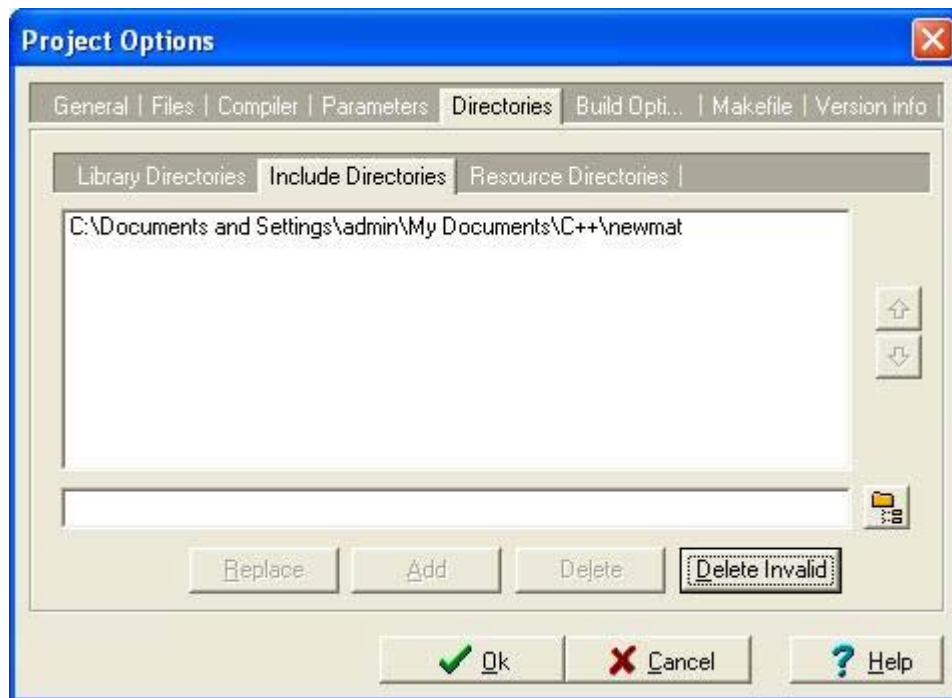


3. In the **Project Options** window, click on the **Directories** tab and do the following.

1. Click on the tab labeled **Include Directories**.
2. Click on the button that looks like a folder with content, like this , in the lower right corner. This pops up a window labeled **Browse for Folder**.



3. In the **Browse for Folder** window, browse to and select the folder `newmat`. We placed the Newmat files in this folder, so it now contains the `newmatio.h` and `newmatap.h` header files that are referenced in the `include` statements in `matrixdemo.h`. Selecting the `newmat` folder will turn it blue; this tells the compiler where to find those header files.
4. Click on the **OK** button; this will dismiss the **Browse for Folder** window, and cause the full path to the `newmat` folder to appear in the **Project Options** window.
5. In the **Project Options** window, click on the button labeled **Add**. This causes an entry for the `newmat` folder to appear in the list under **Include Directories**, in the **Project Options** window. Do not forget to click on the **Add** button, otherwise your folder selection won't "take"! When you're done, the **Project Options** window should now look like this:

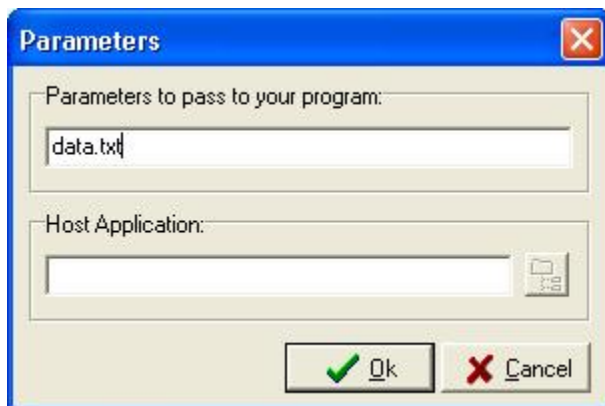


Note that the `newmat` folder is now listed in the large white pane.

6. In the **Project Options** window, click on the **OK** button. This dismisses the **Project Options** window.
7. Select **Execute** → **Parameters....** This opens up a window named **Parameters**. Type the filename

`data.txt`

into the entry field labeled **Parameters to pass to your program:**.



Then click on the **OK** button.

Now that you've done this, when you run the program from within Dev-C++, the name of the data file will be included as the command line argument. It will be as if you had gotten into a DOS window, navigated to the `matrixdemo` folder, and in there typed in the command

```
matrixdemo data.txt
```

I.e., the word "data.txt" will then be passed to the `main` function of the program, and is referred to within the `main` function as `argv[1]` (see Step VI(B) above). If you skip this step, you'll see an error like the following if you attempt to run the program!



8. Press the F9 key to **Compile and Run** the program. If you watch the Status of the compilation closely, you'll notice that there's a distinct Linking step; this is where the `main`, `ReadTextFileIntoMatrix`, `CountWordsInFile`, and `CountLinesInFile` functions as well as the Newmat library all get "linked" together to build the final stand-alone binary executable file, `matrixdemo.exe`.
9. If you get error messages, diagnostic information will be printed out at the bottom of the Dev-C++ window, pinpointing the location (file and line number) of the error. If there are errors, go back to the tabbed editor pane for the file containing the error, fix the error, and then try a **Compile and Run** again. Otherwise, if your code is error-free, you should see the following output displayed in a DOS window.

Least Squares Estimates:

```
Y =
-9.185300
-5.094200
3.127000
7.913400
6.632400
10.097500
19.278500
23.546900
22.957500
26.964900
```

```
X =
1.000000 -9.000000 -2.000000
1.000000 -7.000000 -2.000000
1.000000 -5.000000 3.000000
1.000000 -3.000000 3.000000
1.000000 -1.000000 -2.000000
1.000000 1.000000 -2.000000
1.000000 3.000000 3.000000
```

```

1.000000 5.000000 3.000000
1.000000 7.000000 -2.000000
1.000000 9.000000 -2.000000

```

```

Bhat =
10.623860
2.004162
0.947530

```

Singular Value Decomposition of X:

```

U =
-0.495434 0.258199 -0.316228
-0.385337 0.258199 -0.316228
-0.275241 -0.387298 -0.316228
-0.165145 -0.387298 -0.316228
-0.055048 0.258199 -0.316228
0.055048 0.258199 -0.316228
0.165145 -0.387298 -0.316228
0.275241 -0.387298 -0.316228
0.385337 0.258199 -0.316228
0.495434 0.258199 -0.316228

```

```

D =
18.165902
 7.745967
 3.162278

```

```

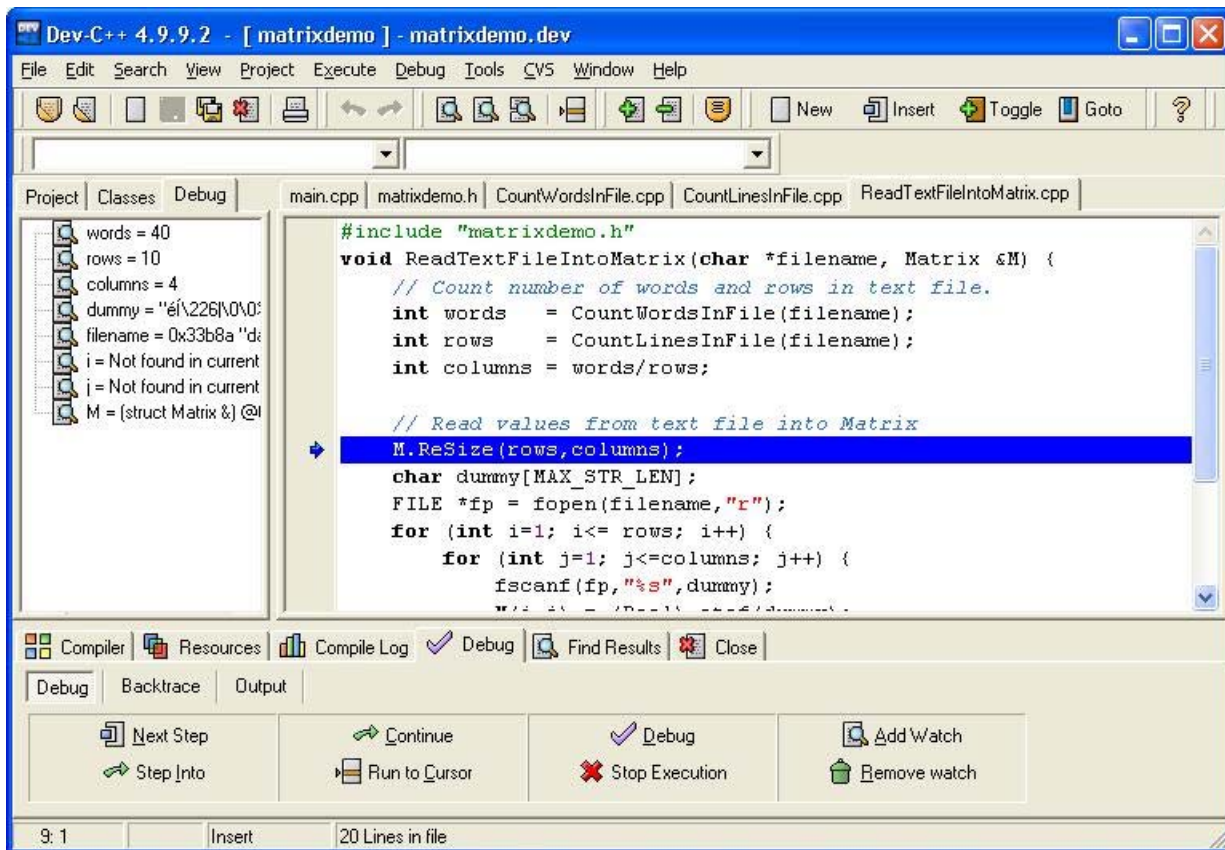
V =
0.000000 0.000000 -1.000000
1.000000 0.000000 0.000000
-0.000000 -1.000000 0.000000

```

Press any key to continue . . .

10. Here's another way to run the compiled program. Get into a Windows file browser and navigate to the `matrixdemo` folder. Click-and-drag the file `data.txt` onto the file `matrixdemo.exe`. You should see the same output that you saw in the previous step.
11. Finally, here's a taste of Dev-C++'s debugger. I must admit that in my hands, the debugger seems a little buggy; sometimes if I did certain things in a certain order, Dev-C++ seemed to hang. Perhaps it is still a work in progress. But let's give this a try anyway. Go to the tabbed editor pane and click on the `ReadTextFilesIntoMatrix.cpp` tab to get to that function. Position the cursor at the very beginning of the 9th line in the file (the line where the Matrix M is resized), and single left-mouse-click there. In the lower left corner of the Dev-C+ window, the numbers "9: 1" should appear, indicating that the editor's cursor is now positioned at the first character of the 9th line.
12. Select **Debug** → **Run to Cursor**; or equivalently, press <Shift-F4>. The DOS window appears as before, but for now contains no text (because we haven't gotten to that part of the program yet). At the same time, the 9th line in `ReadTextFilesIntoMatrix.cpp` is highlighted in blue. This means that the program was executed up to that 9th line, but that execution has now been paused at that line. Notice also that at the same time, the white field on the left has changed; the **Debug** tab is now on top.
- 13.

14. Move your mouse cursor so that it is positioned over any occurrence of the variable `words` in the function, and wait for about two seconds. An entry showing the current value of the variable `rows` should appear in the Debug pane on the left. If not, slightly reposition the cursor over the word `words` and wait two seconds again. Do the same for the `rows`, `columns`, `dummy`, and `filename` variables. Can you get entries for the variables `i` and `j` to appear, too? What about the variable `M`? (I must admit that I wasn't able to get the debugger to display the actual contents of the Matrix `M`; if you're able to figure this out, please email me and tell me how you did it!) The Dev-C++ window will now look something like this:



15. Press the F7 key about twenty times, slowly. Each time you press it, the blue highlight advances one line in `ReadTextFileIntoMatrix.cpp`, and each line is executed in sequence. At the same time, the displayed values of the variables `i`, `j`, and `dummy` are updated in the **Debug** tab.
16. It would be too boring to have to press the F7 key all the way to the end of the `ReadTextFileIntoMatrix` function. So, let's skip all the loops to get to the end of the function. Position the cursor at the very beginning of the 18th line in the file (the line where the `fclose` command appears), and single left-mouse-click there. In the lower left corner of the Dev-C+ window, the numbers "18: 1" should appear. Press <Shift-F4>, so that execution then runs to that line. The `fclose` command is now highlighted in blue, indicating that execution has run up to that point.
17. Press the F7 key twice. Program execution passes back to the `main` function, so the `main.cpp` tab is brought to the front of the tabbed editor pane.
- 18.

19. Press the F4 key; this should bring up a window named **New Variable Watch**. In this window, type

```
(char*)argv[1]
```

so that the window looks like this:



Then press the OK key. This adds the first command line argument to the variables being watched in the **Debug** pane on the left. It should indicate that the current value of `argv[1]` is "data.txt". (A hexadecimal number is also shown before "data.txt" in the **Debug** pane; this is probably the memory address.)

20. Press the F7 key about fifteen times, slowly, and watch the DOS window as execution proceeds through the rest of the program. Text appears in the DOS window as each `cout` command is executed.
21. When you finally get to the `system("PAUSE")` command, you'll have to select the DOS window by clicking on it. Then press any key to continue, as before.
22. Back in the Dev-C++ window, you may need to press the F7 key one or two more times to complete execution of the program, and to make the DOS window finally disappear.

This concludes the tutorial. Of course, we've only barely scratched the surface here. Newmat has many more linear algebra operations for you to use. Similarly, Dev-C++ has many features, and we've only seen the basics. As for C++, there are many books and online tutorials for you to try; I suggest one possible book for further study in the References section below.

6. Suggestions for Further Exploration

1. **Learn more about Newmat's functionality.** I refer you to Dr. Robert Davies' website, where you can find excellent documentation for Newmat. See the References section below.
2. Dr. Davies has his own suggestions on how to compile Newmat using Dev-C++; see http://www.robertnz.net/Dev_C.html. Compare his approach to what we did in Steps V and VI. How do the two approaches differ?
3. **Step VI(G) featured a function `ReadTextFileIntoMatrix`.** Develop a function named `WriteMatrixToTextFile` that does the reverse operation, i.e., that writes a Newmat matrix into a tab- or space-delimited text file, with one row in the text file for every row in the matrix.
4. **Write a program to read in a set of p-values into a Newmat `ColumnVector` from a plain text file.** Sort the p-values from smallest to largest, using the built-in Newmat function for sorting the values in a vector. Then compute a threshold based on False Discovery Rate from the sorted p-values (Benjamini and Hochberg, 1995; Tobia et al., 1999).
5. **Obtain a copy of the famous *Numerical Recipes* book** (the 3rd edition has recently been

published, a milestone in scientific computing. Its name has been truncated from *Numerical Recipes in C*). Using Numerical Recipes code, write software to convert a statistical test (e.g., a t-test or F-test) to a p-value, given the degrees of freedom. Can you figure out how to write software to do the reverse operation: given a p-value and degrees of freedom, compute the corresponding critical threshold of a statistical test?

6. Using what you developed in VII(B) above, extend the simple least-squares estimates program we wrote in step VI to compute p-values.
7. Take all of the *Numerical Recipes* functions and build a statically linked library out of them, like we did in Step V above. Now you have two nifty libraries at your disposal, the Newmat library and the *Numerical Recipes* library!

That's all for this month. If you have any feedback on this column or ideas for future topics, please email me at jmm97@georgetown.edu. As always, your thoughts will be greatly appreciated.

Joe Maisog (again, with thanks to Lanlan Yin for test-driving this tutorial; any errors remain my own)

Georgetown University / Medical Numerics

References.

Benjamini Y and Hochberg Y, Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc Ser B* 57: 289–300 (1995).

Brown D and Satir G, *C++: The Core Language*, Cambridge, MA: O'Reilly, Inc., 1995.

Davies R, Newmat website, <http://www.robertnz.net/nm11.htm> .

Dev-C++ website, <http://www.bloodshed.net/dev/devcpp.html>

Press WH, Teukolsky SA, Vetterling WT, and Flannery BP, *Numerical Recipes 3rd Edition: The Art of Scientific Computing*, 3rd ed., Cambridge: Cambridge University Press, 2007.

Tobia RD, Rom D, Wolfinger RD, Hochberg Y (authors), Westfall PH (ed.), SAS Institute (corporate author), *Multiple Comparisons and Multiple Tests Using the SAS System*, Cary, NC: SAS Institute, 1999.

Employment

As a service to local statisticians, *WSS News* provides notification of employment opportunities and description of those seeking employment here in the Washington, DC, area. Readers are encouraged to take advantage of this feature of the newsletter. The deadline for inserting notices is five (5) weeks before the publication date. Those interested should email or call Anne Peterson, at apeterson@insightpolicyresearch.com or (703) 373-6645.

CLINICAL TRIAL BIOSTATISTICIANS

M.S. and Ph.D. Level Positions

With an opportunity for substantial leadership responsibility in studies of international public health import.

The Biostatistics Center of The George Washington University, founded in 1972, is a leader in the statistical coordination of clinical trials conducted by the National Institutes of Health. We enjoy over \$45 million per year of NIH research funding for major studies in cardiovascular disease, diabetes, maternal/fetal medicine, osteoporosis, urology, and the genetic basis for various diseases. The center has a staff of over 100 with 27 biostatisticians/epidemiologists, including 10 faculty. We are recruiting M.S. and Ph.D. level staff to participate in these and future studies. Please visit our web site (below).

Master's Level Research Positions: These positions require a Master's in Biostatistics or Statistics and 1-5 years experience in analysis, supervision of data management and study design for biomedical applications. Good written and oral communication skills, and detailed knowledge of SAS required. Send CV to address below.

Assistant to Full Research Professorial Positions are available immediately to serve as Co-

Investigator or Principal Investigator (Project Director) and to provide statistical direction of the design, conduct and analysis of studies and the conduct of methodologic research to meet the projects needs. We are seeking individuals who want to join a highly competent team of academic biostatisticians and epidemiologists; who desire to contribute to the design and analysis of major medical studies, seek substantive scientific and statistical responsibility, enjoy interacting with medical investigators; take pride contributing to the publication of major papers in leading medical journals, and desire to make an impact on the public health. Our faculty also participate in graduate programs in biostatistics, epidemiology and statistics which afford opportunities for teaching at the graduate level. The research projects also provide an environment rich in methodological problems, with opportunities for collaboration with research active Center faculty and graduate students.

Minimum Position Requirements: Doctorate in Biostatistics, Statistics or Epidemiology, or alternatively an M.D. or Ph.D. in Biological Science, Physical Science or Computer Science with a Masters in Biostatistics or Statistics, 1-5 years' experience with clinical trials, especially study

design and statistical analysis of study results using SAS, excellent oral and written English communication skills, and supervisory experience.

Application Procedures: Applicants must send a Curriculum Vitae and three letters of reference; a letter to include a synopsis of their role in collaborative medical research that has led to medical scientific presentation or publication and a statement of career purpose indicating their career goals and how this position can help you achieve those goals; and applicants for Assistant Research Professor positions must send an Official Transcript of graduate coursework leading to the doctoral degree to: Sarah Fowler, Research Professor and Director, The George Washington University Biostatistics Center, 6110 Executive Blvd., Suite 750, Rockville, MD 20852.

[HTTP://WWW.BSC.GWU.EDU](http://www.bsc.gwu.edu)

Review of applications is ongoing until the positions are filled. Rank/position title and salary commensurate with experience and qualifications. Tuition benefits for employees (including Ph.D. in Statistics, Biostatistics and Epidemiology) and for spouse and dependent children.

All research and regular faculty at the rank of Assistant Professor in Biostatistics or Statistics may apply for the Samuel W. Greenhouse Biostatistics Research Enhancement Award. For a period of 1 year, the award will provide 20% effort for methodological research and a discretionary fund to support professional activities, travel to professional meetings, supplies and equipment. Applicants for the research faculty position may also apply for the Greenhouse Award while their faculty application is being considered. For complete information including Award Application Materials Requirements, please visit our website at: www.bsc.gwu.edu.

The George Washington University is an Equal Opportunity/Affirmative Action employer

Survey Sampling Statistician

WESTAT: AN EMPLOYEE-OWNED RESEARCH CORPORATION

Westat is an employee-owned corporation headquartered in the suburbs of Washington, DC (Rockville, Maryland). We provide statistical consulting and survey research to the agencies of the U.S. Government and to a broad range of business and institutional clients. With a strong technical and managerial staff and a long record of quality research, Westat has become one of the leading survey research and statistical consulting organizations in the United States.

Our company was founded in 1961 by three statisticians. The current staff of more than 1,800 includes over 60 statisticians, as well as research, technical, and administrative staff. In addition, our professional staff is supported by data collection and processing personnel situated locally and in field sites around the country. The work atmosphere is open, progressive, and highly conducive to professional growth.

Our statistical efforts continue to expand in areas such as the environment, energy, health, education, and human resources. Westat statisticians are actively involved in teaching graduate-level courses in statistical methods and survey methodology in collaborative arrangements with area colleges and universities.

We are currently recruiting for the following statistical position:

Survey Sampling Statistician (Job Code WSS/DRM/7001)

Three or more years of relevant experience in sample design and selection, frames development, weighting, imputation, and variance estimation. Must have a master's or doctoral degree in statistics and have excellent writing skills. Coursework in sample survey design is highly desirable.

Westat offers excellent growth opportunities and an outstanding benefits package including life and health insurance, an Employee Stock Ownership Plan (ESOP), a 401(k) plan, flexible spending accounts, professional development, and tuition assistance. For immediate consideration, please send your cover letter, indicating the Westat Job Code, and resume by one of the following methods to:

Job Code is REQUIRED to apply.

Westat • Attn: Resume System • 1650 Research Boulevard • Rockville, MD 20850-3195

Email: resume@westat.com • FAX: (888) 201-1452

Equal Opportunity Employer. www.westat.com

Chief Mathematical Statistician
Social Security Administration

The Social Security Administration's Office of Research, Evaluation, and Statistics (ORES) is seeking a Chief Mathematical Statistician. In addition to full-time, SSA employment, SSA will consider a detail from other Federal Agencies or an Intergovernmental Personnel Act (IPA) agreement with an academic institution. The Chief Mathematical Statistician reports to the Associate Commissioner and serves as the senior consultant and expert on mathematical and statistical matters. The incumbent recommends and implements statistical policies, assures that appropriate methodologies and practices are used, and serves as ORES' spokesperson on statistical matters.

The Chief Mathematical Statistician must have in-depth, expert knowledge of statistical theory and methods and skill applying this knowledge to highly complex, technical projects, such as:

- Providing advice and serving as lead consultant on the design and use of sample surveys, including sample design, estimation, data analysis, reliability measures, and survey costs/timing.

- Serving as ORES' lead expert and advisor on statistical disclosure avoidance methods, and working with staff to develop, design, and implement current methods.

- Leading the design and implementation of procedures to create public use microdata samples from administrative data.

- Linking data from two independent sources, such as surveys and administrative records, to develop improved estimates.

- Representing ORES at meetings, seminars, and federal statistical working groups.

For more information please contact: Manuel de la Puente, Associate Commissioner, Office of Research, Evaluation, and Statistics at 202/358-6020 or at manuel.de.la.puerta@ssa.gov

Survey Sampling Statistician

The Survey Research Center at ICF International (SRC) is currently seeking a mid-level sampling statistician to conduct sampling, variance estimation and weighting for large, complex survey research projects. The candidate should possess a Master's or doctoral degree in statistics or related discipline. Strong quantitative, writing and communication skills are essential. This position is located in our Fairfax, VA headquarters.

Specific Responsibilities

Construct sample frames for various types of sample surveys
Design and implement sampling, weighting, imputation and variance estimation procedures for large-scale survey projects
Conduct questionnaire design
Conduct analysis of survey data
Prepare project reports
Prepare sampling and analysis sections of technical proposals
Develop written products and discuss issues of sampling, variance estimation and weighting with other team members.

Key Qualifications

Master's or PhD in statistics, biostatistics, or survey methodology

Expert in at least one common statistical language, strong in SAS, familiar with several more statistical languages

Superior general problem solving

Strong communication skills

ICF International offers an excellent benefits package and competitive salary.

To apply, please submit your cover letter and resume to <https://jobs.icfi.com>.

ICF International is an equal opportunity and affirmative action employer.

Researchers

The U.S. Census Bureau is seeking two top-level research experts for Scientific and Professional (ST) Positions:

-Senior Mathematical Statistician to conduct original research in statistics related to censuses and sample surveys.

-Senior Researcher in Survey Methodology to conduct original research in the field of survey methodology, focused on the behavioral and social science aspects of censuses and sample surveys.

Location: Washington, DC Metro area (Suitland, MD)

Salary: \$138,380 - \$158,500

For qualifications, educational requirements, and application procedures:

-Go to <www.usajobs.com>.

-Enter "Senior Mathematical Statistician" or "Senior Researcher in Survey Methodology" in the first search box and "Suitland, MD" in the location search box. (The Senior Mathematical Statistician vacancy number is CEN-08-04P. The Senior Researcher for Survey Methodology vacancy number is CEN-08-02P.)

-Follow the instructions given in the announcement.

Contact Information: Kathy Ott, 301.763.4006, Kathleen.E.Ott@census.gov

Deadline: May 30, 2008

Applications will be accepted from all sources - U.S. Citizens only.

Note from the WSS NEWS Editor

Items for publication in the June issue of the WSS NEWS will be accepted until May 15, 2008. E-mail items to Michael Feil at michael.feil@usda.gov.



P.O. Box 752

Suitland, MD 20752

NON-PROFIT
ORG.
U.S. POSTAGE
PAID

Washington
Statistical
Society

President Michael P. Cohen 202-232-4651	Past President Jill Montaquila 301-517-4046	President-Elect Karol P. Krotki 202-728-2485
Secretary Chris Moriarity 202-327-6414	Treasurer Daniell Toth 202-691-7380	
Vice Chair for District 2 of Council of Chapters Glenn White 202-327-6414	Council of Chapters Representative Jonaki Bose 240-276-1257	
Representative-at-Large Polly Phipps 202-691-7513 Brian A. Harris-Kojetin 202-395-7314	Allen Schirm 202-484-4686 Jeri Mulrow 703-292-4784	
Student Representative Joe Maisog 703-465-2006		
Agriculture and Natural Resources Mel Kollander 703-642-8079 Mike Fleming 202-366-5419 Barry Nussbaum 202-566-1493	Economics Linda Atkinson 202-694-5046 Anne Polivka 202-691-7395	Methodology Meena Khare 301-458-4312 Keith F. Rust 301-251-8278
Public Health and Biostatistics Grant Izmirlian 301-496-7519 Myron Katzoff 301-458-4307	Public Policy Michael L. Cohen 202-334-3765 Shelly Martinez 202-395-3147	Quality Assurance and Physical Sciences Amrut Champaneri 202-366-5998 Alan Jeeves 202-366-4194
Defense and National Security Wendy Martinez 703-696-4320 Myron Katzoff 301-458-4307	Human Rights Wendy Rotz 202-327-7822 Hormuzd Katki 301-594-7818	
Short Courses Sylvia Dohrmann 301-610-5119	Social and Demographic Statistics Judy Droitcour 202-512-9145 Promod Chandbok 202-366-2158	Statistical Computing Charlie Hallahan 202-694-5051 Dean H. Judson 301-763-2057
Electronic Mail Committee Michael Greene 301-504-7335 S.V. (Vince) Massimini 703-983-5893	Data Collection Methods Carol Blumberg 202-586-6565 Eugene Burns 202-586-1385	Membership John Czajka 202-484-4685 John Dixon 202-691-7516 Timothy Kennel 301-763-6795 Adam Safir 202-974-7822 Fritz Scheuren 202-320-3446
Webmaster Dan Jacobs 301-403-4220	WSS NEWS Editor Michael Feil 202-690-3130	Curtis Jacobs Award Tom Krenzke 301-251-4203
Employment Anne Peterson 703-979-1191	Quantitative Literacy Carolyn Carroll 202-320-8709	Local Arrangements Kevin Cecco 202-874-0464
Video Librarian Phil Kalina 703-725-6600	Quantitative Literacy Poster Competition Stacey Chan 202-327-5914 Ryan Petska 202-327-7245	Quantitative Literacy Workshop Mark Otto 301-497-5872
Social Arrangements Yasmin Said 301-538-7478 Yves Thiibaudeau 301-763-1706	WSS Committee on ASA Fellows Steve Cohen 301-427-1466 Dwight Brock 301-517-4026 Carol House 202-720-4333 Fritz Scheuren 202-320-3446	Financial Advisor and Audit Committee Stuart Scott 202-691-7383 R. Clifton Bailey 703-893-8719
Video Coordinator John Dixon 202-691-7516		Historian Tom Mule 301-457-8322
Science Fair Coordinator Bob Clickner 301-294-2815		

Washington Statistical Society Annual Dinner

Wednesday June 25, 2008, 6:30 PM

Cocktails 6:00 PM

Meiwah Restaurant, 4457 Willard Ave, Chevy Chase, MD

(Metro: Friendship Heights on the Red Line)

Keynote Presentation

by

Thomas Lumley, University of Washington

Open Source Statistical Software: Why? When? Where?

- Made Possible by Funding from RTI International -

Cash Bar

Variety of Dishes Including:

Crispy Fried Shredded Beef

And

Chicken with Black Bean Sauce

And

Shanghai Bokchoi with Fresh Mushroom

Price:..... \$45.00 per person

Please register at: <https://www.123signup.com/event?id=tmkbh>

Or call/email: Yves Thibaudeau, (301) 763-1706, Yves.Thibaudeau@census.gov