

WSS NEWS

February 2004



HERRIOT AWARD NOMINATIONS SOUGHT

Nominations are sought for the 2004 Roger Herriot Award for Innovation in Federal Statistics. After the sudden death in May 1994 of Roger Herriot, an Associate Commissioner for Statistical Standards and Methodology at the National Center for Education Statistics, the Washington Statistical Society, the Social Statistics and Government Statistics Sections of the American Statistical Association established an award in his memory to recognize individuals who develop unique approaches to the solution of statistical problems in Federal data collection programs.

The award is intended to reflect the special characteristics that marked Roger Herriot's career.

- * Dedication to the issues of measurement;
- * Improvements in the efficiency of data collection programs; and
- * Improvements and use of statistical data for policy analysis.

The award is not limited to senior members of an organization, nor is it to be considered as a culmination of a long period of service. Individuals at all levels (from entry to senior), Federal employees, private sector employees, and employees of the academic community, may be nominated on the basis of the significance of the specific contribution.

The recipient of the 2004 Roger Herriot Award will be chosen by a committee of representatives of the Social Statistics Section and Government Statistics Section of the American Statistical Association and a representative of the Washington Statistical Society. Roger Herriot was associated with and strongly supportive of these organizations during his career. The award consists of an honorarium of \$1,000.00 and a framed citation.

Joseph Waksberg (Westat), Monroe Sirken (National Center for Health Statistics), Constance Citro (National Academy of Sciences), Roderick Harrison (U.S. Census Bureau), Clyde Tucker (Bureau of Labor Statistics), Thomas Jabine (SSA, EIA, CNSTAT), Donald Dillman (Washington State University), Jeanne Griffith (OMB, NCES, NSF), Daniel Weinberg (U. S. Census Bureau), and David Banks (FDA, BTS, NIST) are previous recipients of the Herriot Award.

For more information, contact Phillip S. Kott, Chair of the Roger Herriot Award Committee, 703-235-5211 x 102 or pkott@nass.usda.gov

Nominations must be submitted by April 15. Electronic submissions, via pdf, are permissible. Alternatively, nominations may be mailed to Phillip S. Kott, USDA/NASS, Room 305, 3251 Old Lee Highway, FAIRFAX, VA 22030-1504.

THE JEANNE E. GRIFFITH MENTORING AWARD

The Jeanne E. Griffith Mentoring Award has been established to encourage mentoring of younger staff in the Federal statistical system. It is presented annually, beginning in 2003, to a supervisor who is nominated by co-workers and supervisors, and chosen by the Award Selection Committee.

The award is co-sponsored by the Interagency Council on Statistical Policy, the Council for Excellence in Government, the Washington Statistical Society, the Social Statistics and Government Statistics Sections of the American Statistical Association, and the Council of Professional Associations on Federal Statistics.

Nominations for 2004 will be accepted beginning in November 2003. The last date for submission of nominations is April 1, 2004, and the Award Committee will make its determination of the award winner by May 1, 2004. The award will consist of a \$1000 honorarium and a citation, which will be presented at a ceremony arranged by the co-sponsors in June 2004. The winning mentor will be selected for his or her efforts in supporting the work and developing the careers of younger staff.

For further information on the award, contact Ed Spar, Council of Professional Associations on Federal Statistics (COPAFS) by phone: 703-836-0404; fax: 703-684-3410; or by e-mail at copafs@aol.com. The nomination cover sheet and guidelines—or a photocopy of it should be attached to a nomination memorandum or letter. Forms can be obtained by contacting Ed Spar, or by downloading from the COPAFS website at http://www.copafs.org. All nominations should be returned to the Jeanne E. Griffith Mentoring Award Committee, c/o COPAFS, 1429 Duke Street, Alexandria, VA 22314 no later than April 1, 2004.

		WSS and Other Seminars (All events are open to any interested persons)
Febru	ary	
11	Wed.	A New Price Index for Air Travel
17	Tues.	Student-t Based Interval Estimation of Complex Statistics Under Calibration Weighting
18	Wed.	Biosurveillance Geoinformatics of Hotspot Detection and Prioritization for Biosecurity
March	า	
3	Wed.	Outliers: Identification and Treatment Through the Use of Chebyshev's Theore

Also available on the World Wide Wed at the following URL: http://www.science.gmu.edu/~wss

Announcements

SIGSTAT Topics for Winter 2004

February 11, 2004: PROC MIXED - Part 4: Random Coefficient Models

Continuing the topic begun in October 2003, after discussing the concepts behind a random coefficient model, PROC MIXED is used to fit such a model. Empirical best linear unbiased predictors (EBLUP's) are then computed. Some common causes of nonconvergence are discussed. Finally, a model with both repeated and random effects is estimated in PROC MIXED.

March 10, 2004: PROC MIXED - Part 5: Model Assessment

Continuing the topic begun in October 2003, we examine residual plots and a histogram of the residuals. Subjects with residual values in the first or last percentiles are identified. Finally, we examine how the violation of assumptions regarding the random effects influences the inference of the model.

April 14, 2004: PROC MIXED - Part 6: Generalized Linear Models & Generalized Linear Mixed Models

Continuing the topic begun in October 2003, the difference between general linear models and models using generalized estimating equations (GEE's) is covered. The available correlation structures in PROC GENMOD are discussed and GENMOD is used to fit a longitudinal data model. Finally, the concepts behind generalized linear mixed models are discussed and a

longitudinal data model is fit using the GLIMMIX macro.

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 12:30 to 1:30. 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 H a I I a h a n, 2 0 2 - 6 9 4 - 5 0 5 1, <u>hallahan@ers.usda.gov</u>, 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/**.

Note from the WSS NEWS Editor

Items for publication in the April 2004 WSS NEWS should be submitted no later than February 24, 2004. E-mail items to Michael Feil at <u>michael.feil@usda.gov.</u>

Title: A New Price Index for Air Travel

Speaker: Janice Lent, Bureau of Transportation Statistics

Discussant: Marshall Reinsdorf, Bureau of Economic Analysis

Date/Time: Wednesday, February 11, 2004, 12:30 - 2:00 p.m.

- Location: Bureau of Labor Statistics, Postal Square Building (PSB), Conference Center, Room 9, 2 Massachusetts Ave., N.W., Washington, D.C. Please use the First Street entrance to the PSB. To gain entrance to BLS, please see notice at the end of this announcement.
- Sponsor: WSS Methodology Section
- Abstract: The Bureau of Transportation Statistics (BTS) is preparing to begin scheduled production of a family of price index series for commercial air travel. The new index series will be based on data from the Passenger Origin and Destination (O&D) Survey, through which BTS collects information from the airlines on a 10% sample of air travel itineraries. Since the Survey was not originally designed to collect data for price index estimation, BTS developed new techniques to estimate Fisher indexes from the O&D Survey data. The large sample allows estimation of index series at geographically detailed levels. We will describe the research performed in developing and testing the new estimation techniques and examine some sample index series computed for research purposes. We will also discuss BTS' future plans for index production and continuous improvement of both the source data and the estimation methods.

* To attend seminars at BLS, you need to email your name, affiliation, and title of the seminar to wss_seminar@bls.gov (underscore between "wss" and "seminar") by noon at least two days in advance, or call 202-691-7524 and leave a message. Bring a photo id to the seminar. A new list begins January 1, 2004. Once you are on the list you need not contact BLS for seminars through March 31, 2004. BLS is located at 2 Massachusetts Ave NE. Take the Red Line to Union Station.

Note: This is the fourth in a series of WSS seminars on calibration and related types of estimation.

- Title: Student-t Based Interval Estimation of Complex Statistics Under Calibration Weighting
- Speaker: Reid A. Rottach, U.S. Census Bureau
- Co-author: David W. Hall, U.S. Census Bureau
- Chair: David W. Hall, U.S. Census Bureau

Date/Time: Tuesday, February 17, 2004, 12:30 - 2:00 p.m.

- Location: Bureau of Labor Statistics, Postal Square Building (PSB), Conference Center, Room 10, 2 Massachusetts Ave., N.W., Washington, D.C. Please use the First Street entrance to the PSB. To gain entrance to BLS, please see notice at the end of this announcement.
- Sponsor: WSS Methodology Section
- Abstract: This seminar gives an overview of recent research for the Survey of Income and Program Participation (SIPP) into developing a linearization variance estimator, including an extension to Satterthwaite-like approximations of degrees of freedom. We provide background about weight calibration, particularly the raking ratio, and the residual technique of estimating variances. Some of the topics covered within this context are nonresponse adjustments, restricted weighting, and weight equalization (such as SIPP's constraint that husbands and wives have the same calibrated weight). A general outline of how to implement the method of constructing confidence intervals is given, along with details for several types of statistics, including totals, ratios, quantiles, and yearly changes. Numerical comparisons with a Balanced Repeated Replication estimator using data from the 1996 panel of SIPP show the two methods of estimating variances to be very close in most cases. Furthermore, we will illustrate circumstances where the degrees of freedom approximation, when compared with the nominal value, substantially affected the confidence interval width.

* To attend seminars at BLS, you need to email your name, affiliation, and title of the seminar to wss_seminar@bls.gov (underscore between "wss" and "seminar") by noon at least two days in advance, or call 202-691-7524 and leave a message. Bring a photo id to the seminar. A new list begins January 1, 2004. Once you are on the list you need not contact BLS for seminars through March 31, 2004. BLS is located at 2 Massachusetts Ave NE. Take the Red Line to Union Station.

Topic: Biosurveillance Geoinformatics of Hotspot Detection and Prioritization for Biosecurity

- Speaker: G. P. Patil, Distinguished Professor and Director, Penn State Center for Statistical Ecology and Environmental Statistics
- Chair: Mel Kollander, Director, Washington Office, Institute for Survey Research of Temple University
- Date/Time: Wednesday, February 18, 2004, 12:30 2:00 p.m.
- Location: Bureau of Labor Statistics, Conference Center Room 9 and 10, Postal Square Building (PSB), 2 Massachusetts Ave. NE, Washington, D.C. Please use the First St., NE, entrance to the PSB. To gain entrance to BLS, please see "Notice" at the end of this announcement.
- Abstract: Geoinformatic surveillance for spatial and temporal hotspot detection and prioritization is a critical need for the 21st century. A hotspot can mean an unusual phenomenon, anomaly, aberration, outbreak, elevated cluster, or critical area. The declared need may be for monitoring, etiology, management, or early warning. The responsible factors may be natural, accidental or intentional, with relevance to both infrastructure and homeland security.

This presentation describes a multi-disciplinary research project based on novel methods and tools for hotspot detection and prioritization, driven by a wide variety of case studies of potential interest to several agencies. These case studies deal with critical societal issues, such as carbon budgets, water resources, ecosystem health, public health, drinking water distribution system, persistent poverty, environmental justice, crop pathogens, invasive species, biosecurity, biosurveillance, remote sensor networks, early warning and homeland security.

Our methodology involves an innovation of the popular circle-based spatial scan statistic methodology. In particular, it employs the notion of an upper level set and is accordingly called the *upper level set scan statistic sytem*, pointing to the next generation of a sophisticated analytical and computational system, effective for the detection of arbitrarily shaped hotspots along spatio-temporal dimensions. We also propose a novel prioritization scheme based on multiple indicator and stakeholder criteria without having to integrate indicators into an index, using Hasse diagrams and partially ordered sets. It is accordingly called *poset prioritization and ranking system*.

We propose a cross-disciplinary collaboration to design and build the prototype system for surveillance infrastructure of hotspot detection and prioritization. The methodological toolbox and the software toolkit developed will support and leverage core missions of several agencies as well as their interactive counterparts in the society. The research advances in the allied sciences and technologies necessary to make such a system work are the thrust of this five year project.

The project will have a dual disciplinary and cross-disciplinary thrust. Dialogues and discussions will be particularly welcome, leading potentially to well considered synergistic case studies. The collaborative case studies are expected to be conceptual, structural, methodological, computational, applicational, developmental, refinemental, validational, and/or visualizational in their individual thrust.

A panel discussion will follow the speaker presentation. Panel participants include: (1) Larry Brandt (NSF), Moderator; (2) Howard Burkom (Johns Hopkins University); (3) Larry Cox (NCHS); (4) John Kelmelis (USGS); (5) Betsy Middleton (NASA); (6) Phil Ross (EPA); (7) Ashbindu Singh (UNEP); and (8) Lance Waller (Emory University). Floor discussion will follow.

NOTE

To be placed on the seminar list at the Bureau of Labor Statistics, e-mail your name, affiliation, and seminar name to <u>wss seminar@bls.gov</u> (underscore after 'wss') at least 2 days in advance of the seminar or call 202-691-7524. Bring a photo ID to the seminar. BLS is located at 2 Massachusetets Ave., NE. Use the Red Line to Union Station.

Title: Outliers: Identification and Treatment Through the Use of Chebyshev's Theorem

Speaker: Richard Esposito, U.S. Bureau of Labor Statistics

Discussant: Hyunshik Lee, Westat

Chair: Fritz Scheuren, National Opinion Research Center

- Date/Time: Wednesday, March 3, 2004, 12:30 2:00 p.m.
- Location: Bureau of Labor Statistics, Postal Square Building (PSB), Conference Center, Room 9, 2 Massachusetts Ave., N.W., Washington, D.C. Please use the First Street entrance to the PSB. To gain entrance to BLS, please see notice at the end of this announcement.
- Sponsor: WSS Methodology Section
- Abstract: Outliers have traditionally been presented as unwanted and troublesome elements whose influence we should protect against through robust methods of data handling and estimation. In this presentation, extreme representative outliers are seen both as inevitable and as providing necessary information about the tail ends of universe distributions. A new method for identifying and treating outliers based on an innovative use of Chebyshev's Theorem is introduced. The method presented avoids the overly wide ranges commonly associated with the use of this theorem, and, being based on Chebyshev's Theorem, does not depend on the normality or specific form of the underlying distribution. The speaker will present data and results demonstrating the importance of outliers, as well as test results that show the ability of the new method to predict corresponding universe values, using universe and drawn samples from national establishment employment data. In addition, a justification and derivation of Winsorizing that follows from considerations of the method will be presented.

* To attend seminars at BLS, you need to email your name, affiliation, and title of the seminar to wss_seminar@bls.gov (underscore between "wss" and "seminar") by noon at least two days in advance, or call 202-691-7524 and leave a message. Bring a photo id to the seminar. A new list begins January 1, 2004. Once you are on the list you need not contact BLS for seminars through March 31, 2004. BLS is located at 2 Massachusetts Ave NE. Take the Red Line to Union Station.

Short Course Announcement

Privacy, Confidentiality, and the Protection of Health Data: A Statistical Perspective on the HIPAA Privacy Rule

A number of developments - requirements of the HIPAA Privacy Rule, NIH requirements for data sharing plans for large grants, together with the proliferation of electronic data and the technical means for intrusions into those data - have intensified the need for guidance from statisticians skilled in data protection.

Over the past several decades, a significant body of research has emerged to form what has come to be known as Statistical Disclosure Limitation Techniques. These methods have not been widely disseminated in the health-related areas and yet are very applicable. The purpose of this Washington Statistical Society (WSS) short course is to provide a basic understanding of the techniques and procedures employed by statisticians and others, to assess disclosure risk in statistical tabulations and electronic data and to minimize this risk to produce analytically useful information. Examples will be drawn from a wide variety of health and medical sources. In addition, the short course will present a review of legal issues related to the confidentiality of health data and, for those cases where a useful data product cannot be safely released or shared, a description of the various means of restricted access currently in use. As part of the course, extensive documentation will be provided to all attendees.

This one-day course, scheduled for April 27, 2004, will cover five areas:

1) Legal issues relating to the confidentiality of health data;

2) Assessing disclosure risk;

3) Statistical disclosure limitation techniques to minimize disclosure risk;

4) Example of disclosure limitation techniques used to protect

health-related data sets; and

5) Restricted access procedures.

About the Speakers:

Donna Eden - Office of General Counsel, Health and Human Services Alvan O. Zarate - Confidentiality Officer, National Center for Health Statistics Philip Steel – Statistical Research Division, Census Bureau Jacob Bournazian - Energy Information Administration, Chair of the Confidentiality and Data Access Committee (CDAC) Steve Cohen - Bureau of Labor Statistics, Vice Chair of the Confidentiality and Data Access Committee (CDAC)

Location (tentative):

BLS Conference Center Postal Square Building 2 Massachusetts Ave., N.E. Washington, D.C. 20212-0001

Please watch this space for complete details and registration information. Interested parties may also contact the short-course co-chairs, to receive registration forms directly:

Sylvia Dohrmann – SylviaDohrmann@westat.com Trena M. Ezzati-Rice - tezzatir@ahrq.gov

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.ipr@verizon.net or (703) 979-1191.

Epidemiologist/GIS Scientist

The National Cancer Institute (NCI) anticipates one opening for a position within the Cancer Statistics Branch (CSB) of the Division of Cancer Control and Population Sciences. This is located within the National Institutes of Health (NIH), Department of Health and Human Services (DHHS). CSB oversees the SEER cancer registries and conducts surveillance studies of health and disease outcomes for various population groups. NCI's surveillance activities provide for nationwide program planning, and are of interest to researchers, policy planners and analysts, and the public at large. Recent examples include: studies of access to cancer care, treatment options and survival, and the role of socioeconomic status in the incidence, mortality, and survival of cancer. Position responsibilities will include: advising on geographic issues related to geographic/spatial data collection, processing and analysis, including efficient geocoding methods, maintaining patient confidentiality in cancer data systems and statistical displays, choice of geographic unit appropriate for statistical analysis, geographic information system (GIS) development, and map design. In addition, opportunities exist for collaboration with SRP staff on studies of small-area population characteristics relevant to cancer screening and health disparities, and statistical, epidemiologic and geovisualization methods useful for the spatio-temporal analysis of cancer data and related demographic and environmental data. Excellent communication skills are necessary to communicate and translate complex information to diverse audiences. Experience with cancer registry data, competing sources of demographic and population data and ESRI software and its extensions is desirable. For more information on Cancer Statistics Branch the see http://surveillance.cancer.gov/csb. For more information on GIS activities at NCI see http://gis.cancer.gov.

A minimum of a Master's degree required in geography, demography, epidemiology, biostatistics or a related field with extensive experience in the application of geospatial techniques to health data. Salary commensurate with experience. The location is Rockville, MD, near Washington, DC. Excellent benefits. DHHS and NIH are equal opportunity employers. Please send a cover letter summarizing your experience and interests along with your CV, and contact information for three references, by March 15, 2004 to: Dr. Benjamin Hankey, Chief, Cancer Statistics Branch, National Cancer Institute, 6116 Executive Blvd., Room 5023, MSC 8316, Bethesda, MD 20892-8316 (US Mail) or Rockville, MD 20852 (Overnight courier); Phone: (301) 496-8510; Fax: (301) 496-9949; Bh43a@nih.gov

Survey Estimation/Experimental Design Research Position

Applications and nominations are invited for a research position in survey estimation and data analysis at the U.S. Census Bureau. The position is in the Statistical Research Division, which has the responsibility for extensive statistical, methodological, and behavioral research relevant to censuses and sample The successful applicant will be surveys. responsible for research, consultation and collaboration in mathematical statistics, with special emphasis on survey design and estimation, nonsampling error modeling, data analysis, and design of experiments. Candidates are expected to demonstrate the potential to make strong theoretical and applied contributions to the Census Bureau's statistical research program.

Qualifications for this position include relevant experience and a record of original research in survey estimation, data analysis or design of experiments, and good communication and interpersonal skills. A Ph.D. in statistics is preferred and experience in computer experiments is desirable. The salary range is \$58,070 to \$89,774, depending on qualifications and experience. U.S. citizenship is required.

For further information or to start the application process, please contact Leroy Bailey, Statistical Estimation and Analysis Research Group, Statistical Research Division, U.S. Census Bureau, Washington, DC 20233 or <u>leroy.bailey@census.gov</u>. You may also call (301) 763-4917.

The Census Bureau Is an Equal Opportunity Employer.

SAS programming positions in Washington DC

Statistics Collaborative, Inc., a small biostatistical consulting firm located in downtown Washington,DC, is currently seeking entry-level and experienced SAS programmers. We are hardworking and committed to high quality, socially important work. Most of our effort deals with clinical trials in a variety of medical fields. Candidates must be literate, numerate, computer savvy and willing to learn new skills. Attention to detail, accuracy, initiative, and excellent communication skills are essential. Knowledge of statistics is desirable, but not required.

Our entry-level programmers have come from a variety of fields, but the most successful ones have been those whose background has been mathematics or science, or, somewhat surprisingly, graphic arts. We want programs that are error-free, well-documented, and that look elegant on the printed page. Salary will depend on your proficiency and expertise.

If you are interested, check our website (www.statcollab.com) and then send all of the following: a cover letter, sample SAS codes you have written, a writing sample, a copy of your academic transcript(s) (unofficial copies are fine), and the names of three references.

Send submissions by email to office@statcollab.com, or by U.S. mail, addressed to:

Ms. Damaya Hicks Statistics Collaborative, Inc. 1710 Rhode Island Ave, NW, Suite 200 Washington, DC 20036



P.O. Box 752 Suitland, MD 20752

Washington Statistical Society NON-PROFIT ORG. U.S. POSTAGE PAID FALLS CHURCH, VA PERMIT NO. 186

POSTMAN: DATED MATERIAL Please deliver by February 1, 2004

President David Marker 301-251-4398		Past President Brenda Cox 703-875-2983		President-Elect John Czajka 202-484-4685	
	Secretar Courtney	y y Stapleton 301-763-4142	Treasur Erin Wh	er itworth Dyal 571-235-7128	
Vice Chair Carolee Bus	for District 2 of Cour h 202-997		Council of Chapters Glenn White	nn White 202-327-6414	
	Dan Cork Tom Mule	Representative-at-Larg 202-334-2573 301-763-8322	ge Jill Montaquila Jennifer Guarino	301-517-4046 301-763-4250	
		WSS	Program Chairs		
Agriculture and Natura Mel Kollaner	al Resources 202-537-6700	Economics Linda Atkinson Anna Jan	202-694-5046 202-327-6328	Methodology Patrick Cantwell Wendy Rotz	301-763-4982 202-327-7822
Public Health and Biost Grant Izmirlian Iai Choi	tatistics 301-496-7519 301-458-4144	Public Policy Brian A. Harris-Kojetin	202-395-7314	Quality Assurance and Amrut Champaneri Eugene Burns	Physical Sciences 202-366-5998 202-366-3491
Short Courses Slyvia Dohrmann Irena Ezzati-Rice	301-610-5119 301-427-1478	Social and Demograph Judy Droitcour	ic Statistics 202-512-9145	Statistical Computing Michael Yang Charlie Hallahan Dean H. Judson	202-223-9160 202-694-5051 301-763-2057
Electronic Mail Commi		Data Collection Methods		Membership	
Michael Greene S.V. (Vince) Massimini	301-504-7335 703- 883-5893	Sameena Salvucci Jonaki Bose	703-807-2309 202-366-9979	Fritz Scheuren John Dixon Timothy Kennel	202-320-3446 202-691-7516 301-736-6795
Webmaster Dan Jacobs	301- 403-4220	WSS NEWS Editor Michael Feil	202-690-3130		
E mployment Anne Peterson	703-979-1191	Quantitative Literacy Carolyn Carroll	703-379-4500	Local Arrangement Kevin Cecco	202-874-0464
V ideo Librarian Mel Kollander	202-537-6700	Quantitative Literacy I Archana Joshee Ryan Petska	Poster Competition 202-327-6859 202-327-7245	Social Arrangements Jeri Mulrow	703-292-4784
Science Fair Coordinate Lee Abramson	or 301-415-6180	WSS Committee on AS Michael L. Cohen Fritz Scheuren	202-334-3765 202-320-3446	WSS Historian & Fina Michael P. Cohen	ncial Advisor 202-366-9949
V ideo Coordinator Iohn Dixon	202-691-7516	Carol House Steve Cohen Dwight Brock	202-720-4333 301-427-1466 301-517-4026	Science Fair Coordinat Lee Abramson	tor 301-415-6180