FEBRUARY 1991

WASHINGTON STATISTICAL SOCIETY

NEWSLETTER

February 19	Tuesday	A New Method for Robust Nonparametric Regression
February 20	Wednesday	The Design and Analysis of Reinterview - An Overview
February 20	Wednesday	Using Cluster Analysis to Identify Health Care Service Areas
February 22	Friday	Spatial Statistics
February 27	Wednesday	Statistical Matching

ANNOUNCEMENTS

Quantitative Literacy Interest Group Request for Volunteers for Montgomery County

An announcement about the WSS Quantitative Literacy (QL) Group appeared in the November 1990 Newsletter. The QL group is looking for volunteers to provide various services in local public schools.

Montgomery County is in the process of revising the math curriculum to incorporate materials from the QL project. They have completed the revisions for grades 7-8 and are now working on grades 9-12. A course entitled "Introduction to Statistics" is offered in grades 10-12.

The Secondary Mathematics Coordinator for Montgomery County, Joy Odom, is particularly interested in having WSS members volunteer to give classroom lectures, provide mentoring for science fair projects, and arrange field trips to places of employment where statistics is practiced. If you are interested, or have any questions, please call the WSS Montgomery County QL Coordinator, Lee Abramson at (301) 492-3949 (work) or (301) 362-7908 (home).

If you are interested in participating in WSS QL activities in other local jurisdictions, please call Ron Fecso at (202) 475-3486 or Dwight Brock at (301) 496-9795.

WASHINGTON STATISTICAL SOCIETY PROGRAM CHAIRS

Agriculture & Natural Resources
Cynthia Clark 763-8558
John Herbert 532-4544

Social & Demographic Statistics
Harvey Schwartz 443-6990
Tom Dietz 323-2916

Short Courses
Glenn White 763-7524
Donald Gantz 764-6565
Brad Pafford 447-3623
Sid Schwartz 268-3490

Economics
John Ruser 523-1347
Neil Ericsson 452-3709

Methodology
Sam Slowinski 452-2622
Sue Ahmed 357-6781

Public Health & Biostatistics
Ed Lakatos 496-5905
Gordon Lan 881-9260

Physical Sciences & Engineering
Nozer Singpurwalla 994-7515
Julia Abrahams 696-4320

Statistical Computing
Nancy Flournoy 885-3127
Sylvia Leaver 272-2350

Quality Assurance
Stanley Freedman 586-2038
John Galvin 272-5066

Newsletter Editor
Stephen H. Cohen 523-7551

Employment
Bill Arends 447-6812

PROGRAM ABSTRACTS

TOPIC: **A NEW METHOD FOR ROBUST NONPARAMETRIC REGRESSION**

SPEAKER: Ferdinand T. Wang, The American University

CHAIR: Nancy Flournoy, The American University

DISCUSSANT: Robert W. Jernigan, The American University

DATE & TIME: Tuesday, February 19, 1991; 12:30 to 2:00 p.m.

LOCATION: Room 2736, GAO Building, 441 G Street, N.W., Washington, D.C. (near the Judiciary Square Metro Station)

SPONSOR: Statistical Computing

ABSTRACT: Consider the problem of estimating the mean function underlying a set of noisy data. Least squares is appropriate if the error distribution of the noise is Gaussian, and if we have good reason to believe that the underlying function has some particular form. But what if the previous two assumptions fail to hold? A statistical procedure is called *robust* if it is relatively insensitive to violations of the assumptions made; in the regression setting, a robust method is one which ignores the outliers caused by an error distribution with heavier tails than the Gaussian. A *nonparametric* regression estimator is one which does not impose the restrictions of any particular family, but rather estimates the function directly from the data. Such methods are appropriate when we do not have any good idea of what form the underlying function takes.

Although we can easily find estimators that are either robust or nonparametric, the literature reveals very few that are both. We propose a new method which uses the fact that the L_1 norm naturally leads to a robust estimator. In spite of the L_1 norm's reputation for being computationally intractable, it turns out that solving the least absolute deviations problem leads to a linear program with special structure. By then utilizing this property over local neighborhoods, we obtain a method that is also nonparametric. Additionally, the new method can generalize to higher dimensions; to date, the problem of smoothing in higher dimensions has met with little success. We present a proof of consistency, find the rate of convergence, and give the results from both simulated and real data.

TOPIC: **THE DESIGN AND ANALYSIS OF REINTERVIEW - AN OVERVIEW**

SPEAKER: Irwin Schreiner, Census Bureau

CHAIR: Ron Fecso, NASS

DISCUSSANT: Dale Atkinson, NASS

DATE & TIME: Wednesday, February 20, 1991; 12:30 to 2:00 p.m.

LOCATION: Room 5152, Agriculture South Building, 14th and Independence, S.W., Washington, D.C.

SPONSOR: Methodology Section

ABSTRACT: Reinterview is a powerful tool for controlling field work and estimating measurement errors in surveys. The method was developed in the United States and India during the 1940s and has been used in a number of countries since then. In this paper we discuss the different aspects of the reinterview. This includes the definition(s), purposes, design, development of the reinterview questionnaire, reconciliation, time lag, field implementation issues, evaluating interviewer performance, use of computer assisted telephone and personal interviewing, and survey error models for analyzing reinterview data.

PROGRAM ABSTRACTS (continued)

TOPIC: **USING CLUSTER ANALYSIS TO IDENTIFY HEALTH CARE SERVICE AREAS**

SPEAKER: Diane M. Makuc, Chief, Analytical Coordination Branch, Division of Analysis, NCHS

CHAIR: Joel C. Kleinman, Director, Division of Analysis, NCHS

DISCUSSANT: Thomas Moore, Chief, Surveys and Supplements Branch, Statistical Methods Division, Bureau of the Census

DATE & TIME: Wednesday, February 20, 1991; 2:00 to 3:30 p.m.

LOCATION: National Center for Health Statistics, Auditorium, Presidential Building, 11th Floor, 6525 Belcrest Road, Hyattsville, Maryland 20782

SPONSORS: Office of Research and Methodology, NCHS and the Washington Statistical Society

ABSTRACT: This presentation describes a methodology to identify health care service areas for the United States. This project was motivated by a decision at NCHS to consider health care service areas as possible primary sampling units (PSUs) for the National Health Interview Survey and other NCHS surveys that use the same PSUs as the NHIS. A health care service area is defined as a county or group of counties that are relatively self-contained with respect to the provision of routine hospital care. The distance between any two counties is defined as a function of the ratio of flow of patients between two counties to the number of hospital stays in the smaller county. 1988 Medicare data on short-stay hospital stays were used to calculate the distance between counties. Nonhierarchical cluster analysis and the average linkage method were used to group counties, based on this distance measure. Four alternative solutions were generated: 800 areas with no constraints on MSAs (800U), 800 areas with counties in an MSA constrained to be within the same service area (800L), 1400 areas with MSAs unconstrained, and 1400 areas with MSA counties linked. All four solutions produced service areas that are adequately self-contained for the majority of the U.S. population. However, the 800 area solutions are more self-contained in the less populated areas of the country. The 800U and 800L solutions differ in the following respects: the 800U areas split 63 MSAs into 2 or more service areas, are more likely to link a nonmetro county with a metro county, and have a more uniform distribution of number of counties per area.

TOPIC: **SPATIAL STATISTICS**

SPEAKER: Paul Switzer, Stanford University

DATE & TIME: Friday, February 22, 1991; 11:00 a.m.

LOCATION: Lecture Room A, National Institute of Standards and Technology (Route 270 North to Clopper Road exit; lecture room is in Administration Building).

SPONSORS: Physical Sciences and Engineering Section, WSS, and the Statistical Engineering Division, NIST

ABSTRACT: Dr. Switzer will present a personal view of spatial statistics. The emphasis will be on the role of spatial data in building models of stochastic fields and the role of stochastic field models in assessing uncertainty of spatial interpolation.

PROGRAM ABSTRACTS (continued)

TOPIC: STATISTICAL MATCHING

SPEAKERS: Mary Ceist, IRS
Michael Cohen, University of Maryland/Brookings Institution

CHAIR: Constance Citro, NAS

DISCUSSANT: Dan Kasprzyk, Census Bureau

DATE & TIME: Wednesday, February 27, 1991; 12:30 to 2:00 p.m.

LOCATION: Room 2736, GAO Building, 441 G Street, N.W., Washington, D.C.

SPONSOR: WSS Methodology Section

ABSTRACT: Michael Cohen will speak on Statistical Matching and Microsimulation Models: a Review." Microsimulation models are typically large, computationally demanding models which make use of data at the level of the decision-making unit in order to determine the impact of program changes. One problem typically faced is that the information needed for input into a micro-simulation model is not available on any one data file. As a result, statistical matching, a type of merging of data files, is used to create a file composed of records with complete information. The well-known assumptions underlying statistical matching are reviewed, along with some methods for reducing the dependence on these assumptions. Finally, characteristics of statistical matching particular to its use in microsimulation models are discussed.

Mary Geist will report on research done by Mary Geist and Fritz Scheuren on "An Analysis of an Underlying Assumption for Statistical Matching." Statistical matching is a technique in which each record from one data source is matched with a record from a second source which does not represent the same unit. An underlying assumption in this technique is statistical independence between the matched variables given the variables in common. This presentation investigates the validity of this assumption. Measures of association using Cross-Product ratios have been compared between unmatched and matched data sets. These sets include the Survey of Income and Program Participation (unmatched data) and a Statistics of Income-Current Population Survey file (matched data).

ANNOUNCEMENTS (continued)

Tentative Schedule of SIGSTAT Meetings

SIGSTAT is the Special Interest Group in Statistics in the Capital PC User Group. The tentative schedule of events through April is as follows:

2/13/91	Derive - symbolic math package.
3/13/91	Forecast Master - time series forecasting.
4/10/91	Shazam - a very complete econometric estimation package.
5/15/91	???? wildcard - suggestions to Charlie Hallahan.

All meetings are scheduled for Wednesdays from 12:30 to 1:30 p.m. in Room B-14, 1301 New York Avenue, N.W. The building is located midway between the Metro Center and McPherson Square Metro stops. If this is your first SIGSTAT meeting, call Charlie Hallahan, 786-1507, and leave your name in order to gain entry into the building.

EMPLOYMENT COLUMN

The Washington Statistical Society Newsletter provides a service of notification of employment opportunities and descriptions of those seeking employment here in Washington. Readers are encouraged to take advantage of this feature of the newsletter. Deadline for inserting notices is 5 (five) weeks before the publication date. Those interested should write to: Bill Arends, USDA-NASS, Room 4133 South Building, Washington, D.C. 20250-2000, Phone 447-6812.

JOB OPENINGS

U.S. Department of Agriculture
National Agricultural Statistics Service

The National Agricultural Statistics Service, the primary collector of agricultural data in the United States, has openings for mathematical statisticians and related disciplines at various entry levels. Salaries are commensurate with qualifications and experience. Persons with a B.S., M.S., or Ph.D. in a relevant field, or equivalent experience, are encouraged to apply. Positions involve both applied and methodological research work in such areas as:

- Sample design, estimation and variance estimation
- Control and evaluation of nonsampling error
- Statistical methods for analyzing survey data
- Analysis of multivariate earth resource observation satellite data
- Measurement error models
- Statistical graphics and statistical computing
- Total survey error and cost models
- Crop yield forecasting models
- New computer-assisted data collection and processing technologies

These positions are central to the primary mission of NASS to measure agricultural output and production of the food and fiber sector of the U.S. economy. Persons interested in these positions should send letters of interest and resumes to Dr. Cynthia Clark, Room 4801 South Building, National Agricultural Statistics Service, USDA, Washington, D.C. 20250-2000. An Equal Opportunity Employer, U.S. citizenship required.

STATISTICIAN: GS 11/12

Research Associate with skills in research design, statistical analysis, large-scale database management, and data processing using SAS and other mainframe microcomputer software packages is wanted to participate in planning and conducting mental health studies of the causes and effects of traumatic events. Experience with interpretation of survey and experimental research findings, and with presentation of statistical results in tabular and graphic format is essential. Also desirable is experience in conducting literature surveys and familiarity with recent hardware advances. Salary range $31,116 to $48,481 depending on experience. Send resume to: Susan D. Solomon, Ph.D., Violence and Traumatic Stress Research Branch, National Institute of Mental Health, 5600 Fishers Lane, Room 18-105, Rockville, MD 20857. For more information, contact 301/443-3728.

JOB OPENINGS (continued)

SURVEY/HEALTH STATISTICIANS

(Service Fellow Positions-equivalent to GS 11-14)

The Agency for Health Care Policy and Research (AHCPR) is recruiting for statisticians to join the Division of Statistics and Research Methodology within the Center for General Health Services Intramural Research (CGHSIR). Positions are available for a Ph.D. in statistics/biostatistics and a M.S. with experience in sample design, survey research, sampling, weights development, data analysis for complex surveys, imputation procedures and matching techniques. Familiarity with statistical software packages (particularly SAS) is required. Send SF-171 application forms and resume to: Dr. Steven B. Cohen, Director, Division of Statistics and Research Methodology, Agency for Health Care Policy and Research, Room 18A-55, 5600 Fishers Lane, Rockville, Maryland 20857.

STATISTICIAN, GS-13

The Petroleum Marketing Division of the Energy Information Administration is currently recruiting a GS-13 mathematical statistician. Incumbent is responsible for developing and evaluating sampling frames, sample designs, estimation techniques, and imputation procedures. Incumbent also evaluates and validates survey data and analyzes discontinuities caused by changes in sample designs or the rotation of the sample. Send an SF-171 and a copy of your latest performance appraisal to Paula Weir, EI-431, 1000 Independence Avenue, S.W., Washington, D.C. 20585.

VISITING FACULTY POSITIONS 1991 - 1992

The Department of Statistics/Computer and Information Systems, of the George Washington University, has openings for one or two one-year appointments as Visiting Assistant/Associate Professor, in the area of statistics or statistical computing, beginning Fall 1991. The positions require a Ph.D. in the relevant area. Applicants for these positions should have a strong commitment to both teaching and research in statistics, statistical computing, or computer science. Applicants from all areas of statistics are invited.

Send applications including curriculum vitae, relevant reprints and three letters of reference to Professor A. D. Kirsch, Chairman, Department of Statistics/Computer and Information Systems, The George Washington University, Washington, D.C. 20052 by March 15, 1991. George Washington University is an equal opportunity/affirmative action employer.

P.O. Box 70843
Washington, D.C. 20024-0843