Index

Chapter 1

ASA Bayesian (3 + 1 Competition)

Program Chair: **Stangl, Dalene K.**

1.1 asa.bayesian.01

Sponsoring Section/Society: ASA-SBSS

Session Slot: 8:30-10:20 Monday
Estimated Audience Size: 100
AudioVisual Request: Overhead Projector, Slide Projector

Session Title: Bayesian Analysis and Public and Private Policy Making

Theme Session: Yes
Applied Session: No

Session Organizer: **Zellner, Arnold** University of Chicago

Address: Arnold Zellner Grad. Sch. of Business U. of Chicago 1101 E. 58 St. Chicago, IL 60637
Phone: 773-702-7145
Fax: 773-702-0458
Email: arnold.zellner@gsb.uchicago.edu

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 0 minutes First Speaker - 30 minutes Second Speaker - 30 minutes Third Speaker - 30 minutes First Discussant - 10 minutes Second Discussant - 10 minutes Floor Discusion - 0 minutes

Session Chair: **Zellner, Arnold** University of Chicago

Address: Arnold Zellner Graduate School of Business U. of Chicago 1101 E. 58 St. Chicago, IL 60637
Phone: 773-702-7145
Fax: 773-702-0458
Email: arnold.zellner@gsb.uchicago.edu

1. Mutual and Pension Funds Management: Beating the Markets using a Global Bayesian Investment Strategy

Quintana, Jose Mario, CDC, Investments, New York

Address: Jose Mario Quintana CDC Investment Management Corp. 9 West 57th St., New York, NY 10019
Phone: 212-891-1823
Fax: 212-891-6113
Email: quintana@cdcna.cdcc.com
Putnam, Bluford H., CDC, Investments, New York

Abstract: *There is a trend in mutual and pension funds investments toward indexing (i.e., to buy and hold "the market"). This tendency is fueled by the apparent confirmation of the efficient market hypothesis given the inability of the majority of active U.S. equity managers to beat the S&P 500 index in recent years. However, the concept of an efficient market is a relative one. In this presentation we will show a global Bayesian strategy that can generate attractive risk-adjusted excess returns by exploiting inefficiencies in the broad global markets. Furthermore, these excess returns are transportable, allowing the strategy to outperform any index, including top-performing indices such as the S&P 500.*

2. What to Do When the Crystal Ball is Cloudy: Conditional and Unconditional Forecasting in Iowa

Whiteman, Charles, University of Iowa

Address: Charles H. Whiteman Pioneer Hi-Bred Professor of Financial Economics Chair, Department of Economics W210 PBAB The University of Iowa Iowa City, IA 52242

Phone: 319-335-0831

Fax: 319-335-1956

Email: whiteman@uiowa.edu

Otrok, Christopher, University of Iowa

Abstract: Since 1990, the predictions of state tax revenues in Iowa which provide the foundation for official forecasts have been made using Bayesian methods under asymmetric linear loss. The procedures utilize vector autoregressions and uninformative priors; predictive distributions are calculated using the Monte Carlo method. Unconditional forecasts have been remarkably accurate. During policy discussions surrounding the 1997 income tax cut, conditional forecasts utilizing the methods were used to assess the usefulness of outside consultants' "structural" model estimates of the effects of the cut. This effort suggested that the uncertainty associated with the structural, behavioral effects of the tax cut was small relative to the inherent "reduced form" uncertainty in revenue forecasting.

3. Why use hierarchical models to assess medical profiles?

Morris, Carl, Harvard University

Address: Carl Morris Dept. of Statistics Harvard U. Cambridge, MA 02138

Phone: 617-4965-1602

Fax: 617-495-8057

Email: morris@stat.harvard.edu

Christiansen, Cindy L., Harvard University

Abstract: Hierarchical models are being used for quality assessment more and more frequently, including for profiling medical units. In what sense are they better than simpler methods, and can we decide before fitting them whether they will produce noticeably different results? What pitfalls must be watched for, and how? Do we need to describe multi-level models and present their results more simply in order to widen their acceptance for regulatory use? The talk addresses these questions, partly through examples involving hospital comparisons.

Discussant: **Polson, Nicholas** University of Chicago

Address: Nicholas Polson Grad. Sch. of Business Univ. of Chicago 1101 E. 58th St. Chicago, IL 60637
Phone: 773-702-9298
Fax: 773-702-0458
Email: nicholas.polson@gsb.uchicago.edu

Discussant: **Dorfman, Jeffrey** University of Georgia

Address: Jeffrey H. Dorfman Associate Professor of Ag. & Applied Economics 315 Conner Hall University of Georgia Athens, GA 30602-7509
Phone: 706-542-0754
Fax: 706-542-0739
Email: jdorfman@agecon.uga.edu
List of speakers who are nonmembers: None

1.2 asa.bayesian.02

Sponsoring Section/Society: ASA-SBSS

Session Slot: 10:30-12:20 Tuesday
Estimated Audience Size: 80-120
AudioVisual Request: Overhead Projector, Slide Projector

Session Title: Inference for Deterministic Models in Environmental Policy

Theme Session: Yes

Applied Session: No

Session Organizer: **Raftery, Adrian E.** University of Washington

Address: Adrian Raftery Department of Statistics University of Washington Box 354322 Seattle, WA 98195-4322
Phone: 206-543-4505
Fax: 206-685-7419
Email: raftery@stat.washington.edu

Session Timing: 110 minutes total (Sorry about format):

> Opening Remarks by Chair - 5 First Speaker - 30 minutes Second Speaker - 30 minutes Third Speaker - 30 minutes Discussant - 0 minutes Floor Discusion - 15 minutes

Session Chair: **Raftery, Adrian E.** University of Washington

Address: Adrian Raftery Department of Statistics University of Washington Box 354322 Seattle, WA 98195-4322
Phone: 206-543-4505
Fax: 206-685-7419
Email: raftery@stat.washington.edu

1. Assessing Uncertainty in Projections of Forest Growth Under Varying Climate Scenario

Green, Edwin J., Rutgers University

Address: Prof. Edwin J. Green Dept of Ecology, Evolution & Natural Resources Cook College - Rutgers University New Brunswick, NJ 08903-0231
Phone: 732-932-9152
Fax: 732-932-8746
Email: green@crssa.rutgers.edu
Valentine, Harry T., USDA Forest Service

Abstract: Valentine et al. have developed a mechanistic forest growth model, called PIPESTEM. The model has been calibrated for loblolly pine. When used in conjunction with the determinsitc *carbon flux model MAESTRO, PIPESTEM can deliver predictions regarding the growth of loblolly pine under various scenarios regarding climate change and atmospheric carbon dioxide concentrations. Unfortunately, the projections suffer from the fact that they are not accompanied by any statement of uncertainty. We present our work on determining uncertainty estimates for loblolly pine growth projections under specified climatic scenarios, using the Bayesian Synthesis method to incorporate incorporate expert knowledge and auuxilliary data.*

2. ***Supra-Bayesian Pooling for Simulation Models***

Roback, Paul, Colorado State University

Address: Paul Roback Department of Statistics Colorado State University Fort Collins, CO 80523
Phone: 970-491-3824
Fax: 970-491-7895
Email: proback@stat.colostate.edu
Givens, Geoff, Colorado State University

Abstract: We consider a deterministic simulation model, such as phi=M(theta). For instance, theta could be whale mortality and productivity, while phi could be whale population and age distribution; the mapping M could be invertible, or it could be non-invertible and highly complex. Our goal is to make Bayesian inference about either the inputs (theta) or the outputs (phi) when prior distributions (representing expert opinion) and likelihoods (based on data which has been collected) are available for both inputs and outputs. At the heart of this problem, expert opinions which have been rendered in different probability spaces must be combined into one probability space. One method for combining these opinions is the supra-Bayesian approach, in which the experts' prior opinions are considered data for a meta-decision maker, who uses Bayes' Theorem to update his/her beliefs. Advocates of this method tout its ability to account for miscalibration, dishonesty, and nonindependence of the experts, as well as its adherence to several desirable philosophical principles. We describe

the supra-Bayesian approach and its motivations, and we apply it in several examples.

3. Bayesian Synthesis Inference for Noninvertible Deterministic Simulation Models via Geometric Pooling

Poole, David, University of Washington

Address: David Poole Department of Statistics University of Washington Box 354322 Seattle, WA 98195-4322

Phone: 206-543-8484

Fax: 206-685-7419

Email: poole@stat.washington.edu

Raftery, Adrian E., University of Washington

Abstract: Two key issues with the Bayesian synthesis assessment method are considered. The first is the Borel paradox, according to which different parameterizations of model inputs and outputs lead to different results. A modification of the Bayesian synthesis method called geometric pooling, proposed last year by Raftery, Poole and Givens (1996), is extended so as to apply to inputs in noninvertible models. It is applied to a simplified population dynamics model (PDM) for bowhead whales. This turns the Bayesian synthesis method into a fully Bayesian method and so avoids the Borel paradox.

The second issue is that of relabelling inputs as outputs or vice versa. For example, in the "forwards" variant, initial population size is considered as an input and current population size as an output, while in the "backwards" variant it is the other way round. The two variants yield different results. Recognizing that the deterministic PDM is used only as an approximation to a more realistic stochastic PDM, we extend the Bayesian synthesis method to stochastic mechanistic models. We find that the forwards variant gives results that are close to those from a stochastic PDM, while the backwards variant gives results that are quite different. We also provide a mathematical explanation of the difference between the forwards and backwards variants, and propose a further modification, the "full pooling" method, that resolves it.

List of speakers who are nonmembers: None

1.3 asa.bayesian.03

Sponsoring Section/Society: ASA-SBSS

Session Slot: 2:00- 3:50 Wednesday
Estimated Audience Size: 100
AudioVisual Request: none

Session Title: Bayesian Methods in Health-Care Policy

Theme Session: Yes
Applied Session: No

Session Organizer: **Stangl, Dalene** Duke University

Address: Dalene Stangl Box 90251, ISDS Duke University Durham, NC 27708-0251
Phone: 919-684-4263
Fax: 919-684-8594
Email: dalene@stat.duke.edu

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 5 minutes First Speaker - 30 minutes Second Speaker - 30 minutes Third Speaker - 30 minutes Discussant - 10 minutes Floor Discusion - 5 minutes

Session Chair: **Sargent, Dan** Mayo Clinic

Address: Dan Sargent Cancer Center Statistics Mayo Clinic 200 1st St. SW Rochester, MN 55905
Phone: 507-284-5380
Fax: 507-284-1902
Email: Sargent.Daniel@mayo.edu

1. Stochastic Optimization Methods For Cost-Effective Quality Assessment in Health Care

Draper, David, University of Bath, UK

Address: David Draper Statistics Group School of Mathematical Sciences University of BAth Claverton Down Bath BA2 7AY, England

Phone: +44-1225-826222

Fax: +44-1225-826492

Email: d.draper@math.bath.ac.uk

Abstract: *Government health services worldwide have recently been looking more closely at the quality of care they provide, focusing in particular on what goes on inside hospitals. One method for indirectly assessing the quality of hospital care is a kind of input-output (I/O) approach, in which hospital outcomes are compared after adjusting for differences in inputs. With death as the monitored outcome, the input-output approach takes the form of a contrast between observed and expected mortality rates, given how sick patients are when they arrive at the hospital. A broad implementation of I/O hospital quality assessment would involve data-gathering on many thousands of patients per year, making the cost-effective measurement of admission sickness crucial to the success of health policy initiatives of this kind.*

Clinical expert judgment typically identifies about 100 variables relevant to the construction of a sickness scale for each disease. The standard way to construct such a scale uses logistic regression with death as the outcome, sifting through available sickness indicators with standard variable-selection methods to find a parsimonious and clinically reasonable subset. From a cost-effectiveness point of view this approach is deficient in that it takes no account of differences in cost of data collection among the available predictors. When both data-collection cost and accuracy of prediction of mortality are considered, a large optimization problem arises in which costly variables that do not predict well enough should be omitted from the scale.

In this paper I take a Bayesian decision-theoretic approach (based on maximization of expected utility) to solving this optimization problem. I compare the usefulness of several methods of stochastic optimization based on MCMC, including simulated annealing and simulated tempering, and competitors including genetic algorithms. I illustrate the methods with data on 2700 el-

derly pneumonia patients from a large nationally-representative study of Medicare patients hospitalized in the US between 1980 and 1986.

2. Evaluating profiling methods used in the health care industry to identify extreme performance

Christiansen, Cindy L., Harvard Medical School and Harvard Pilgrim Health Care

Address: Cindy L. Christiansen, Ph.D. Assistant Professor of Ambulatory Care & Prevention Harvard Medical School & Harvard Pilgrim Health Care 126 Brookline Ave., Suite 200 Boston, MA 02215
Phone: 617-421-6013
Fax: 617-421-2716
Email: cindy@hphc.org

Abstract: Health care profiles are performed and reported routinely by government, the media, and by private organizations. Often, profile reports are used to identify extreme performance, e.g., which providers have extremely high or extremely low utilization, patient satisfaction, or patient outcome rates. Health care statisticians recently have suggested and used hierarchical models to improve the accuracy of these profile reports and to enhance the value of information for decision makers and health care consumers. We use simulation studies to compare hierarchical modeling methods for detecting outliers with commonly-used exact and approximate p-value methods. These evaluations are made for several settings which mimic the characteristics of actual data.

3. Conditional Categorial Response Models with Application to Treatment of Acute Myocardial Infarction

Gelfand, Alan E., Univ. of Connecticut

Address: Alan E. Gelfand Department of Statistics, U-120P University of Connecticut Storrs, Connecticut 06269-3120
Phone: 860-486-3416

Fax:
Email: GELFAND@UCONNVM.UCONN.EDU

Abstract: *For a sample of 2361 patients admitted with suspected acute myocardial infarction (AMI) to a set of 37 hospitals, recorded patient response variables inc lude eligibility for treatment with aspirin, eligibility for treatment with thrombolytics, treatment with aspirin received, treatment with thrombolytics received, and short- term patient survival. Each of these five variables has two levels resulting in a* 2^5 *contingency table. Covariate information includes age, sex, race and comorbidity status. Because the responses arrive in sequence we model this data in three st ages: eligibility, then treatment received given eligibility and finally short-term survival given eligibility and treatment received, all given the covariates. Issues of i nterest include extent to which treatment received matches eligibility, whether pobability of survival is affected by treatment status and how the chance of mortality is affected by whether or not treatment received matches eligibility. Influence of covariate information on these quantities is examined. These quantities are studi ed at the hospital level adjusted for case mix and also in aggregate, marginalizing over hospitals.*

Discussant: **Stangl, Dalene** Duke University

Address: Dalene Stangl Box 90251, ISDS Duke University Durham, NC 27708-0251
Phone: 919-684-4263
Fax: 919-684-8594
Email: dalene@stat.duke.edu
List of speakers who are nonmembers: None

1.4 asa.bayesian.04

Sponsoring Section/Society: ASA-SBSS

Session Slot: 10:30-12:20 Thursday
Estimated Audience Size: 100
AudioVisual Request: Overhead Projector, Slide Projector

Session Title: Bayesian Generalized Linear Models

The role of the statistician in public policy is to help stakeholders effectively use information and accurately quantify uncertainty when making decisions. This often requires blending of information from many samples or sites. This blending creates several problems: 1) the increased heterogeniety observed across samples is not accounted for using conventional models, 2) geographic correlation of samples may need to be incorporated, and 3) estimates may be needed at both local and global levels. Bayesian hierarchical models are especially well suited to solving these problem. In this session speakers will discuss how Bayesian Generalized Linear Models can be used to solve these problems using examples from environmental monitoring, geographic and temporal disease mapping, and risk assessment.

Theme Session: No
Applied Session: No

Session Organizer: **Sun, Dongchu** University of Missouri

Address: Dongchu Sun Department of Statistics University of Missouri Columbia, MO 65211
Phone: 573-882-7675
Fax: 573-884-5524
Email: dsun@stat.missouri.edu

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 5 First Speaker - 30 minutes Second Speaker - 30 minutes Third Speaker - 30 minutes Discussant - 10 minutes Floor Discusion - 5 minutes

Session Chair: **Speckman, Paul L.** University of Missouri-Columbia

Address: Paul L. Speckman Department of Statistics 222 Math Sciences Building University of Missouri-Columbia Columbia, MO 65211

Phone: 860-486-3414
Fax: 860-486-4113
Email: dey@stat.uconn.edu

1. Overdispersed Generalized Linear Models

Dey, Dipak K., University of Connecticut

Address: Dipak K. Day Department of Statistics University of Connecticut Storrs, CT 06269-3120
Phone: 860-486-3414
Fax: 860-486-4113
Email: dey@stat.uconn.edu

Abstract: Generalized linear models have become a standard class of models for data analysts. However in some applications, heterogeneity in samples is too great to be explained by the simple variance function implicit in such models. A two parameter exponential family which is overdispersed relative to a specified one parameter exponential family yields a the class of overdispersed generalized linear models (OGLM's) which are analytically attractive. We propose fitting such models within a Bayesian framework employing noninformative priors in order to let the data drive the inference. Hence our analysis approximates likelihood-based inference but provides an entire posterior distribution for model parameters. Bayesian calculations are carried out using a Metropolis sampling algorithm. An example involving damage incidents to cargo ships motivates our work. Details of the data analysis are provided including comparison with the standard generalized linear models analysis. Several diagnostic tools reveal the improved performance of the OGLM.

2. Hierarchical and Empirical Bayes Methods for Environmental Risk Assessment

Sankar, Gauri Sankar, University of Georgia

Address: Gauri Sankar Datta Department of Statistics University of Georgia Athens, GA 30602
Phone: 301-457-4728

Fax: 301-457-2299
Email: gauri@stat.uga.edu
Ghosh, Malay, University of Florida
Waller, Lance A., University of Minnesota

Abstract: The cancer and other disease atlases have become important tools for environmental monitoring and risk assessment. To study geographical variation in incidence of disease and mortality rates for local areas preparation of these maps has received considerable attention in recent years. Such maps usually display either relative rates in each area, as measured by a standardized mortality ratio (SMR), or p-values based on some statistical tests. None of these methods is entirely satisfactory. The raw SMR estimates are often based on small sample sizes, and hence are usually unreliable. In order to borrow strength from the local areas and to smooth the raw estimates a model-based procedure is often appropriate. We will review the recent developments in hierarchical Bayes generalized linear models including random effects for local areas. Both hierarchical and empirical Bayes estimation will be considered.

3. Default Bayesian Analysis for Generalized Linear Model

Sun, Dongchu, University of Missouri

Address: Dongchu Sun Department of Statistics University of Missouri Columbia, MO 65211
Phone: 573-882-7675
Fax: 573-884-5524
Email: dsun@stat.missouri.edu
Tsutakawa, Robert, University of Missouri

Abstract: Bayesian analysis for a generalized linear model, where the prior follows a hierarchical linear mixed model, has received much attention recently. In practice, it is often very difficult to choose the hyperparameters. This paper first gives necessary and sufficient conditions for the propriety of the posterior

distribution in hierarchical linear mixed effects models for a collection of improper prior distributions. In addition to the flat prior for the fixed effects, the collection includes various limiting forms of the invariant gamma distribution for the variance components, including cases considered previously by Datta and Ghosh (1991), Datta (1992), and Hobert and Casella (1996). Previous work is extended by considering a family of correlated random effects which include as special cases the CAR (1) models by Besag, York and Mollie (1991) and Clayton and Kaldor (1987) and autoregressive model by Ord (1975), which have been found useful in the analysis of spatial effects. A noninformative prior for the spatial correlation coefficients are also examined. Conditions are then presented for the propriety of the posterior distribution for generalized linear mixed models. The methods are successfully applied for estimating the mortality or disease rates including demographic variables such as age and sex, random geographic variables such as regional differences, longitudinal variables such as temporal trends in mortality, and spatial variables such as the correlation between neighboring regions.

Discussant: **Albert, Jim** Bowling Green State University

Address: Jim Albert Department of Mathematics and Statistics Bowling Green State University Bowling Green, OH 43403
Phone: 419-372-7456
Fax: 419-372-6092
Email: albert@math.bgsu.edu
List of speakers who are nonmembers: None

Chapter 2

ASA Biometrics (4 + 2 competition)

Program Chair: **D'Agostino, Sr, Ralph**

2.1 asa.biometrics.01

Sponsoring Section/Society: ASA-BIOMETRICS

Session Slot: 2:00- 3:50 Tuesday
Estimated Audience Size: 100-200
AudioVisual Request: one projection equipment for using laptop, others overhead projector

Session Title: Effects of Smoothing on Health Effects Maps

The National Center for Health Statistics (NCHS) has published and the National Cancer Institute (NCI) will soon publish mortality atlases in which maps by cause, race, and sex depict geographic patterns of disease mortality. The NCHS atlas maps both raw and smoothed rates (age-adjusted) , while it was decided that the NCI atlas will not include smoothed rates for several reasons, one of which is the fear of masking true "hot spots" caused by localized risk factors.

In light of the opposing decisions by NCHS and NCI on inclusion of smoothed rates, I am organizing an ASA session for 1998 entitled "Effects of Smoothing on Health Effects Maps." A rule-of-thumb subscribed to by many is that smoothing is helpful for viewing geographical patterns, but is not appropriate when local variation is important to monitor. The session would focus on elucidating appropriate methods and uses of smoothing for health effects maps and for comparing smoothed and unsmoothed procedures for computing rates in terms of performance criteria.

Theme Session: Yes
Applied Session: Yes

Session Organizer: **Pennello, Gene** National Cancer Institute DCEG/BB

Address: Gene Pennello National Cancer Institute DCEG/BB Executive Plaza North, Room 403 6130 Executive Blvd MSC 7368 Bethesda MD 20892-7368
Phone: (301)496-3344
Fax: (301)402-0081
Email: pennellg@epndce.nci.nih.gov

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 5 or 0 minutes First Speaker - 30 minutes (or 25) Second Speaker - 30 minutes Third Speaker - 30 minutes Discussant - 10 minutes (or none) Floor Discussion - 10 minutes (or 5 or 15)

Session Chair: **Pickle, Linda W.** National Center for Health Statistics

Address: National Center for Health Statistics 6525 Belcrest Road Rm 915 Hyattsville MD 20782
Phone: (301)436-7904 ext 148
Fax: (301)436-7955
Email: lwp0@cdc.gov

1. Bayesian Inference For Extreme Rate From Health Effects Maps

Cressie, Noel, Iowa State University

Address: 102F Snedecor Hall Department of Statistics Iowa State University Ames, IA 50011-1210
Phone: 515-294-3441
Fax: 515-294-4040
Email: ncressie@iastate.edu
Stern, Hal S., Iowa State University

Abstract: *Health effects (e.g., disease-incidence rates) are often reported for small areas and displayed using a choropleth map, such as one might find in cancer atlases. Areas with extremely high incidence rates may be subject to further scrutiny in an attempt to identify possible risk factors. Hierarchical probability models for disease-incidence rates can be used to obtain the joint posterior distribution for an entire set of small-area rates while taking into account the spatial nature of the data. Optimal estimators of small-area incidence rates will differ depending on whether ensemble estimation is the goal or special emphasis is placed on other functions, such as extreme values (e.g., optimal estimation of the top ten rates and the small areas corresponding to them). A loss function for extremes will be developed and MCMC methodology will be used to carry out the subsequent Bayesian inference. The results will be compared to a simpler approach obtained from constrained Bayesian ensemble estimation.*

2. Ranks of Extreme Rates by Summary Rates For Geographic Health Data

Pennello, Gene, National Cancer Institute DCEG/BB

Address: Gene Pennello National Cancer Institute DCEG/BB Executive Plaza North, Room 403 6130 Executive Blvd MSC 7368 Bethesda MD 20892-7368
Phone: (301)496-3344
Fax: (301)402-0081
Email: pennellg@epndce.nci.nih.gov

Abstract: *Methods for summarizing disease rates by geographic area for purposes of mapping include the use of crude* *rates, age-standardized rates (direct and indirect), and empirical Bayes rates. Epidemiologists focus on the highest summary rates because they can suggest leads to disease etiology. This note uses simulation to compare summary rates by their ranking of the areas with the highest underlying risks of disease. Sensitivities and specificities at ranking the designated areas above and below cut points in rank are computed and used to produce receiver operating characteristic (ROC) curves. These ROC curves are compared with the mean square errors of the summary rates decomposed into the variance of the summary rate and the bias introduced by confounding age effects. Data were simulated assuming the Poisson-Gamma mixed model for disease counts and relative risks with parameters based on cancer mortality data for blacks and whites from the National Center for Health Statistics. The results were used in deciding on the method of choice for the maps in the upcoming atlas of cancer mortality from the National Cancer Institute.*

3. Effects of Smoothing Mortality Data Using the Weighted Head-Banging Algorithm

Mungiole, Michael, National Center for Health Statistics

Address: National Center for Health Statistics 6525 Belcrest Rd., room 915 Hyattsville, MD 20782
Phone: 301-436-7904, x145
Fax: 301-436-7955
Email: mim4@cdc.gov
Simonson, Katherine Hansen, Sandia National Laboratories
Pickle, Linda W., National Center for Health Statistics

Abstract: Maps of smoothed data permit the reader to identify general spatial trends by removing the "background noise" of random error in the original data. We used the weighted head-banging algorithm (Mungiole et al. 1996) on mortality data to consider the effects of varying smoothing parameters in an attempt to obtain a more objective measure for determining the appropriate degree of smoothing. We also modified the smoothing algorithm to consider the effects of increasing the number of *nearest neighbors being considered when smoothing areas along the perimeter of the contiguous United States. As was found in previous smoothing research with mortality data sets, weighted headbanging smoothed spikes and edge effects in the perimeter area data. Increasing the number of nearest neighbors in the smoothing process resulted in moderate changes to the areas along the perimeter, with the border more closely resembling areas in close proximity to it. When varying the degree of smoothing, it was found that the appropriate degree of smoothing was somewhat dependent on the amount of noise in the original data. The process used to obtain the smoothed data, including the choice of headbanging parameters, is discussed.*

4. Should Spatial Maps Be Smoothed?

Simon, Gary, New York University

Address: Department of Statistics and Operations Research Stern School of Business, New York University 44 West Fourth Street New York NY 10012-1126

Phone: 212-998-0451

Fax: 212-995-4003

Email: gsimon@stern.nyu.edu

Abstract: *The values on a spatial map are smoothed for at least two reasons, improved estimation and enhancement of visual effect. The improved estimation logic simply exploits local sameness and then averages to produce estimates of lower variance, much as one might do to a time series periodogram. We will consider the use of smoothing for visual effect. It will be argued that, on substantive issues alone, some maps should not be smoothed. Further, we will propose measures to assess visual smoothness and develop the statistical properties of these measures. It will be claimed that smoothing should not be done on maps for which the measures do not reach statistical significance. For those maps on which the measures do reach significance, it will be possibly to assess quantitatively the improvement in visual smoothness.*

List of speakers who are nonmembers: None

2.2 asa.biometrics.02

Sponsoring Section/Society: ASA-BIOMETRICS

Session Slot: 2:00- 3:50 Monday
Estimated Audience Size: 250-300
AudioVisual Request: Overhead Projector, 2*2 slide projector

Session Title: Estimating the Cumulative Risk (Penetrance) Associated with a Mutation

Disease-associated mutations have been and are being identified, often through studies of selected families with many affected members. To understand the public health importance of these mutations, and to provide informed counselling for people carrying such mutations, there is a need to estimate their associated disease risk (penetrance) in the general population. The speakers will discuss the concept of penetrance and compare the strengths and weaknesses of various designs to estimate penetrance, including standard epidemiologic designs, such as cohort and case-control studies, and designs based on representative nuclear families.

Theme Session: No
Applied Session: Yes

Session Organizer: **Gail, Mitchell** National Cancer Institute

Address: Mitchell Gail, M.D., Ph.D. National Cancer Institute 6130 Executive Blvd, EPN431 Rockville, MD 20892
Phone: 301-496-4156
Fax: 301-402-0081
Email: gailm@epndce.nci.nih.gov

Session Timing: 110 minutes total (Sorry about format):

110 minutes total...please allocate Opening Remarks by Chair - 5 or 0 minutes First Speaker - 30 minutes (or 25) Second Speaker - 30 minutes Third Speaker - 30 minutes Discussant - 10 minutes (or none) Floor Discussion - 10 minutes (or 5 or 15)

Session Chair: **Olson, Jane M.** Case Western Reserve University

Address: Department of Epidemiology and Biostatistics Case Western Reserve University MetroHealth Medical Center 2500 MetroHealth Drive Cleveland, Ohio 44109-1998
Phone: (216) 778-4057
Fax: (216) 778-3280
Email: olson@darwin.cwru.edu

1. Penetrance – What Is It?
Wijsman, Ellen, University of Washington

Address: Div. of Medical Genetics and Dept. Biostatistics BOX 357720, University of Washington Seattle, WA 98195-7720
Phone: (206) 543-8987
Fax: (206) 616-1973
Email: wijsman@u.washington.edu

Abstract: Rapid advances in human genetics are leading to identification of many genes involved in diseases of interest in public health. One use of this new information is to quantify the risks of disease to relatives or to estimate the fraction of cases in the population which are attributable to mutations in these genes. This requires estimation of parameters for the penetrance function, which gives the relationship between the observed phenotype, y, and the underlying genotype, g, possibly also including environmental covariates, c. The penetrance function, P(y—g,c) for a discrete trait, with an anlagous function for a continuous trait, is based on the implicit assumption that it represents an average over a randomly selected set of individuals with genotype g. However, because of the rarity of most disease mutations, studies aimed at estimating parameters for the function rarely use randomly selected individuals; the sampling framework typically involves pedigrees, clinic populations, or other high-density populations of individuals. This, coupled with various biological complexities which we often fail to take into account in defining our functions, can lead to great disparity among penetrance estimates obtained in different studies. I will use a number of real *data examples to illustrate some of the issues which influence estimates of penetrance, including effects of modifier loci, different mutations, ascertainment biases, and even how we assay the phenotype of interest.*

2. A Comparision of Cohort, Case-Control and Family Study Designs for Estimating Penetrance

Gail, Mitchell, National Cancer Institute

Address: National Cancer Institute 6130 Executive Blvd, EPN431 Rockville, MD 20892

Phone: (301)-496-4156

Fax: (301)-402-0081

Email: gailm@epndce.nci.nih.gov

Abstract: One can obtain population-based estimates of the penetrance of a measurable mutation from cohort studies, from population-based case-control studies, and from "genotyped-proband designs" (GPD). In a GPD, representative individuals(probands) agree to be genotyped, and then one obtains information on the phenotypes of first degree relatives. We also consider an extension of the GPD in which a relative is genotyped(the GPDR design). We compare the sample sizes needed for these various designs to yield required precision for penetrance estimates. We also discuss the feasibility and validity of these various designs and mention that the GPD and GPDR designs are more susceptible to a bias that results when the tendency for an individual to volunteer to be a proband or to be a subject in a cohort or case-control study depends on the phenotypes of his or her relatives.

3. Estimating Disease Risk From Family-Based Case-Control Studies and Improving Penetrance Estimates With Hierarchical Modeling

Witte, John, Case Western Reserve University

Address: John S. Witte, PhD Department of Epidemiology and Biostatistics Case Western Reserve University 2500 MetroHealth Drive Cleveland, OH 44109-1998

Phone: (216) 778-8523
Fax: (216) 778-3280
Email: witte@darwin.cwru.edu

Abstract: *The case-control design offers an attractive approach for estimating disease risk due to candidate genes and gene-environment interactions. The conventional approach of using population controls allows one to estimate population-based penetrance. When using a population-based design, however, potential problems of population stratification must be addressed. To avoid this issue one can instead select family members as controls. Here, I quantify the efficiency of using various family-based designs to estimate risk. Regardless of the design used, standard analytic techniques will often give unstable and unreasonable penetrance estimates due to the large number of relatively rare mutations in disease genes. Hence, I will also discuss how one can attempt to improve standard estimates of penetrance with hierarchical modeling.*

Discussant: **Schaid, Daniel J.** Mayo Clinic

Address: Daniel J. Schaid, Ph.D. Harwick 7 Mayo Clinic Rochester, MN 55905
Phone: 507-284-0639
Fax: 507-284-9542
Email: schaid@mayo.edu
List of speakers who are nonmembers: None

2.3 asa.biometrics.03

Sponsoring Section/Society: ASA BIOMETRICS

Session Slot: 8:30-10:20 Wednesday
Estimated Audience Size: 400-500
AudioVisual Request: Overhead and slides

Session Title: Statistics and Policy Setting at the Food and Drug Administration

Theme Session: Yes
Applied Session: Yes

Session Organizer: **Ellenberg, Susan S.** Center for Biologics Evaluation and Research, FDA

Address: Center for Biologics Evaluation and Research, FDA 1401 Rockville Pike, HFM-210 Rockville, MD 20852-1448
Phone: (301)-827-3034
Fax: (301)-827-3529
Email: ellenberg@cber.fda.gov

Session Timing: 110 minutes total (Sorry about format):

110 minutes total Opening Remarks by Chair - 5 or 0 minutes First Speaker - 30 minutes (or 25) Second Speaker - 30 minutes Third Speaker - 30 minutes Discussant - 10 minutes (or none) Floor Discussion - 10 minutes (or 5 or 15)

Session Chair: **Cambell, Gregory** Center for Devices and Radiological Health, FDA

Address: Director, Division of Biostatistics Center for Devices and Radiological Health, FDA 1350 Piccard Drive, HFZ-542 Rockville, MD 20850
Phone: (301)-594-0616
Fax: (301)-443-8559
Email: campbell@cdrh.fda.gov

1. Statistics, Statisticians and Policymaking at the FDA

Ellenberg, Susan S., Center for Biologics Evaluation and Research, FDA

Address: Center for Biologics Evaluation and Research 1401 Rockville Pike, HFM-210 Rockville MD 20852-1448
Phone: (301)-827-3034
Fax: (301)-827-3529
Email: ellenberg@cber.fda.gov

Abstract: *Statisticians at the Food and Drug Administration are intimately involved with the development and evaluation of new medical products. They assess study designs and data bases from preclinical investigations through the definitive clinical trials. This is perhaps the most well-understood role of statisticians at the FDA. But much effort in any regulatory agency goes into the development (and revision) of regulations, guidance documents, and other policy development activities. Statisticians contribute to such efforts in many ways. Some contributions relate to technical statistical issues, such as setting thresholds for establishing the purity or potency of a biological product on the basis of a laboratory assay. Others relate to broader issues, where statistics is one of several disciplines relevant to the policy issue in question. Examples would be policies on collecting and evaluating data on age and gender effects of new medical interventions; on use of appropriate control groups for clinical trials of new pharmaceuticals; and on approaches to ensuring the integrity of data submitted in support of marketing applications. Statisticians with long experience at the FDA or in medical research generally will frequently be part of policy teams in areas that are not specifically statistical at all, but that require in-depth knowledge of issues in drug development and/or clinical research.*

2. ***Active Control Equivalence Trials in Drug Development***
Temple, Robert J., Center for Drug Evaluation and Research, FDA

Address: Center for Drug Evaluation and Research Food and Drug Administration 5600 Fishers Lane, HFD-101 Rockville, MD 20857-001
Phone: (301)-594-6758
Fax: (301)-594-5298
Email: temple@cder.fda.gov

Abstract:

3. ***The Role of the Statistician on FDA Advisory Committees***
D'Agostino, Sr., Ralph, Boston University

Address: Statistics and Consulting Unit Department of Mathematics Boston University 111 Cummington Street Boston, MA 02215

Phone: (617)-353-2767

Fax: (617)-353-8100

Email: ralph@math.bu.edu

Abstract: Each FDA Advisory Committee has a statistician member, who can play a major role in the committee's deliberations and ultimately on the decision and policies of the FDA divisions. The statistician is asked repeatedly to be the primary evaluator of the design, execution and analysis aspects of clinical trials, observational studies and post-marketing surveys. With appropriate diligence to the substantive questions under discussion, including the statistical and non-statistical aspects, the statistician can be extremely influential. He/she can bring a new level of rigor to the Advisory Committee. Often the committee meetings offer an ideal setting for promulgating statistical concepts into the medical arena. In this talk I present reflections on my experience serving on advisory commitees, experience that started in 1974. I will use topics including study design, study population, interim analysis, longitudinal analysis, analysis of cluster data, multiple comparison, multiple testing, meta-analysis and label comprehension as examples to illustrate where, I believe, advisory committee statisticians have made tremendous impact. I will also discuss aspects of a statistician that are instrumental in being a successful advisory committee statistician.

Discussant: **O'Neill, Robert T.** Center for Drug Evaluation and Research, FDA

Address: Center for Drug Evaluation and Research, FDA 5600 Fishers Lane, HFD-700 Rockville, MD 2085a7-001

Phone: (301)-827-3195

Fax: (301)-480-2825

Email: oneill@cder.fda.gov

List of speakers who are nonmembers: Temple, Robert J.

2.4 asa.biometrics.04

Sponsoring Section/Society: ASA-BIOMETRICS

Session Slot: 8:30-10:20 Tuesday
Estimated Audience Size: 150-200
AudioVisual Request: LCD Projection Panel (SVGA capable) with connection for laptop,Overhead, slide projector

Session Title: Issues in Using Regression, Classification Trees, and Neural Networks for Predicting Binary Outcomes

This session explores the use of regression models, classification trees, and neural networks for the construction of prediction models for binary outcomes. Over the past twenty years statistical prediction models for binary outcomes have become widely used to assess and predict the probability of an event. For example, regression models are central to health policy and health services research, and particularly over the past decade, are used as clinical decision aids. Logistic regression models for medical events are central to most current probabilistic predictive clinical decision aids and severity of illness tools, and are fundamental to comparative analyses of medical care based on risk-adjusted events. This session will look at the utility and limitations of three widely used methods for constructing these prediction models.

Theme Session: No
Applied Session: Yes

Session Organizer: **Ruthazer, Robin** New England Medical Center

Address: Robin Ruthazer, MPH New England Medical Center 750 Washington Street, NEMC #63 Boston, MA. 02111
Phone: (617)-636-8819
Fax: (617)-636-5560
Email: rruthazer@es.nemc.org

Session Timing: 110 minutes total (Sorry about format):

110 minutes total...please allocate Opening Remarks by Chair - 5 or 0 minutes First Speaker - 30 minutes (or 25) Second Speaker

- 30 minutes Third Speaker - 30 minutes Discussant - 10 minutes (or none) Floor Discussion - 10 minutes (or 5 or 15)

Session Chair: **Schmid, Christopher** Tufts University School of Medicine

Address: Tufts University School of Medicine New England Medical Center 750 Washington Street, NEMC #63 Boston, MA. 02111
Phone: (617)-636-5179
Fax: (617)-636-8023
Email: cschmid@es.nemc.org

1. A Comparision of Regression Models, Calssification Trees, and Neural Networks for the Prediction of Cardiac Complications

Griffith, John, New England Medical Center

Address: Tufts University School of Medicine Director Biostatistics Research Center New England Medical Center 750 Washinton Street, Nemc #63 Boston, MA. 02111
Phone: (617)-636-4619
Fax: (617)-636-5560
Email: John.griffith@es.nemc.org
Ruthazer, Robin, New England Medical Center
Schmid, Christopher, Tufts University School of Medicine
Terrin, Norma, New England Medical Center

Abstract: Prediction models of various types have long been central to health policy and health services research, and particularly over the past decade, as clinical decision aids. We report results of a comparison of models constructed using regression methods, classification trees, and neural networks to predict cardiac complications among patients presenting to the emergency department with ischemic symptoms. Data used were collected in a large prospective multi-center clinical trial of 10,783 patients. Models were constructed on a random subset of these patients and performance tested on the remaining data. Modeling methods included standard logistic regression, regression models with *smoothing splines, generalized additive models, recursive partitioning with pruning and shrinkage, and back-propagated neural networks. On the independent test dataset the accuracy of model performance and predictions, including the bias, area under the ROC curve, calibration chi-square, mean discrimination, and ratio of number of cases in highest to lowest quintile of predictions were compared.*

2. Logistic Regression - Tools for Measuring and Improving Predictive Accuracy

Lee, Kerry L., Duke University Medical Center

Address: Community and Family Medicine Duke University Medical Center P.O. Box 3363 Durham, NC 27710
Phone: (919)-286-8725
Fax: (919)-286-2947
Email: lee0001@mc.duke.edu
Harrell, Frank, Duke University Medical Center

Abstract: *Logistic regression, along with other multivariable regression models, is widely used in risk assessment and in studies of clinical outcomes. Uncritical application of modeling techniques, however, can result in models that inadequately fit the dataset at hand, or even more likely, inaccurately predict outcomes for new subjects. It is important to measure qualities of a model's fit in order to avoid poorly fitted or overfitted models. This talk will discuss several measures for quantifying predictive accuracy, including an easily interpretable measure of predictive discrimination as well as methods for assessing calibration of predicted probabilities of the outcome event. Both types of predictive accuracy should be unbiasedly validated using bootstrapping or cross validation before applying a model to a new group of subjects. Using real-life examples with actual clinical data and readily available software tools, the talk will also present some strategies for avoiding the hazards of poorly fitted or overfitted regression models and the tendency to place too much trust in the unvalidated fit of a model.*

3. Hybrid CART-Logit for Classification and Segmentation

Steinberg, Dan, Salford Systems

Address: Salford Systems 8800 Rio San Diego Dr. San Diego, CA 92108
Phone: (619)-543-8880
Fax: (619)-543
Email: dstein@salford-systems.com
Cardell, Scott N., Salford Systems

Abstract: CART and logistic regression are among the most used classification and response probability modeling tools and both have exhibited from good to excellent performance in a variety of data analysis problems. Since the two methods have quite different strengths and weaknesses, it is natural to investigate whether some combination might prove superior to either used separately. We introduce a new method for combining CART and logit which does exhibit performance superiority, and which admits of a natural set of statistical tests for the incremental value of the combination. This method differs considerably from previous hybridization experiments in that the logistic component is not run within CART child nodes; thus our hybrid exploits CART's ability to detect local data structure, and uses the logit to recognize global structure.

4. Neural Networks: Advantages and Limitations for Statistical Modeling

Goodman, Philip H., University of Nevada School of Medicine, Washoe Medical Center

Address: Division of General Internal Medicine Department of Internal Medicine University of Nevada School of Medicine Washoe Medical Center 77 Pringle Way Reno, NV 89520
Phone: (702)-328-4869
Fax: (702)-328-4871
Email: goodman@unr.edu
Harrell, Frank, Duke University Medical Center

Abstract: The principal advantage of an artificial neural network (ANN) is that a single probabilistic architecture can simultaneously and flexibly capture the effects of predictor curvilinearity and global interaction. Theoretically, then, a (properly fitting) ANN is more likely than a GLM to be a "true" model of a complex predictive relationship. As a screening tool, an ANN can efficiently indicate when evidence of generalizable nonlinearity is lacking, thereby justifying the use of only a simple GLM (decreasing both the risk of finding false positive associations, and the investment of human effort). A general approach to improving predictive modeling through the complementary use of ANNs and GLMs will be demonstrated through the analyses of several large datasets. The presentation will focus on issues of overfitting and estimating optimistic bias in calibration and discrimination, on the interpretation of nonlinear predictive "effects", and on Bayesian evidence-based hyperpenalties for variable selection. Free software and tutorials may be downloaded from ftp.scs.unr.edu/pub/cbmr/nevpropdir.

List of speakers who are nonmembers: None

2.5 asa.biometrics.05

Sponsoring Section/Society: ASA Biometrics

Session Slot: 2:00- 3:50 Sunday
Estimated Audience Size: 100-200
AudioVisual Request: overhead and slides projector

Session Title: Spatio-Temporal Modelling of Small Area Health Data

This session focusses on the a developing area within epidemiology which has considerable import for informing policy in health resource allocation and assessment of environmental health risk. The analysis of spatio-temporal patterns of disease incidence has considerable importance in areas of health surveillance(early detection of adverse incidence) and in understanding of epidemiological processes. This session features a number of diverse approaches to this field which is fast evolving, and a spectrum of aopplication areas is also represented,from disease cluster detection to the analysis of hidden structure via mixtures.

Theme Session: Yes
Applied Session: Yes

Session Organizer: **Lawson, Andrew** University of Abertay Dundee

Address: DR Andrew B. Lawson Mathematical Sciences Division University of Abertay Dundee Bell Street Dundee DD1 1HG UK
Phone: 44-1382-308604
Fax: 44-1382-308627
Email: a.lawson@tay.ac.uk

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 0 minutes First Speaker - 25 minutes Second Speaker - 25 minutes Third Speaker - 25 minutes Fourth Speaker - 25 minutes Floor Discusion - 10 minutes

Session Chair: **D'Agostino, Ralph B.** Boston University

Address: Statistics and Consulting Unit, Department of Mathematics Boston University 111 Cummington Street Boston, MA 02215
Phone: 617-353-2767
Fax: 617-353-8100
Email: ralph@math.bu.edu

1. Space-Time Mixture Modelling in Health Data

Boehning, Dankmar, Free University, Berlin

Address: PD Dr Dankmar Bohning Department of Epidemiology Institute of Social Medicine Free University Berlin Haus 562 Fabeckstr 60-62 14195 Berlin Germany
Phone: 49-30-8445-1292
Fax: 49-30-8445-1290
Email: boehning@zedat.fu-berlin.de
Schlattmann, Peter, Free University Berlin

Abstract: *Mixture models for the construction of disease maps have been suggested in parametric(Clayton and Kaldor, Biometrics 1987, 671-687) and nonparametric form (Schlattmann and Boehning, Statistics in Medicine 1993,1943-1950). The discrete nature of the nonparametric mixture model estimator has been exploited to bind each component with a colour (or grey pattern) in the map which leads to a natural way of disease map construction with the program DISMAP=20(Schlattmann, 1996, Statistics in Medicine, 931). However, in these models time effects are not allowed. Here we consider ways to incorporate replications of area counts of disease cases in the mixture model which then would enable the disesase-mapper to model space-time clustered data simultaneously.*

2. On Space-Time Cluster Modelling via MCMC Methods

Lawson, Andrew, University of Abertay Dundee, UK

Address: Mathmatical Science Division University of Abertary Dundee Bell Street Dundee DD1 1HG UK
Phone: 44-1382-308604
Fax: 44-1382-308627
Email: a.lawson@tay.ac.uk
Clark, Allan, University of Abertay

Abstract: *The analysis of disease clustering is of considerable importance when time of occurence as well as spatial location are observed. This is especially true when surveillance systems are to be established. In this work we adopt a parametric modelling approach where clusters are regarded as objects to be estimated and where different types of clusters can be defined,depending on their space-time characteristics. The development of MCMC methods based on reversible jump sampling is extended to the case where temporal and spatial and spatio-temporal clusters are allowed.*

3. On the Bias of the Knox Method and Other Space-Time Interaction Tests

Kulldorff, Martin, National Cancer Institute

Address: Dr Martin Kulldorff Biometry Branch DCPC National Cancer Institute Bethesda Maryland 20892-7354 USA
Phone: 301-496-7519
Fax: 301-402-8816
Email: martink@helix.nih.gov
Hjalmas, Ulf, Ostersund Hospital, Sweden

Abstract: *The Knox method, as well as other tests for space-time interaction, are biased when there are geographical population shifts so that different regions have different percentage population growth. In this talk, the size of the population shift bias is investigated for the Knox test, and it is shown that it can be a considerable problem. A method for constructing unbiased space-time interaction tests is then presented and illustrated. Practical implications are discussed in terms of the interpretation of past results and the design of future studies.*

4. Spatio-Temporal Modelling of Mortality Data for Multiple Causes of Death

Leyland, Alistair, University of Glasgow, UK

Address: Dr Alastair Leyland Public Health Research Unit University of Glasgow 1 Lilybank Gardens Glasgow G12 8RZ UK
Phone: 44-141-330-5399
Fax: 44-141-337-2776
Email: a.leyland@udcf.gla.ac.uk
Langford, Ian, University of East Anglia
Goldstein, Harvey, University of London
Rashbash, J., University of London

Abstract: *This paper illustrates how it is possible to develop a spatial model for counts of diseases occurring in discrete geographical areas as a multilevel model with heterogeneity effects and spatial effects considered to be random and correlated across areas. Such an approach has the advantage that it permits the extension of the model to include other levels of nesting. The model*

can therefore be extended to include temporal effects either by considering the time points to be repeated measures within areas or by modelling random variation as a function of temporal distance across areas in the same way that physical distance is modelled for the spatial effects. Similarly, a further extension means that multiple responses can be modelled for each area within each time point. This approach is illustrated on cause-specific mortality data for small geographical area population approximately 1000) within Scotland.

List of speakers who are nonmembers: None ???

2.6 asa.biometrics.06

Sponsoring Section/Society: ASA Biometrics Section

Session Slot: 10:30-12:20 Thursday
Estimated Audience Size: 300-350
AudioVisual Request: Overhead

Session Title: The Impact of Patient Compliance

The need to assess patient compliance in clinical trials involving medication is widely recognized. The failure of patients to adhere adequately to prescribed medication is an important medical problem. Studies have demonstrated the prevalence of poor adherence across all types of regimens and diseases. Treatment efficacy cannot adequately be evaluated without valid measures of patient adherence. The use of careful and sensitive instruments to measure adherence behavior in a clinical trial is important in developing new therapies.

Unless appropriate account is taken of compliance problems in clinical trials, policies based on these trial findings may be misguided. Moreover, overlooking compliance problems in the general population may yield health care policy recommendations that are impractical or ineffective in actual application.

Theme Session: No

Applied Session: Yes

Session Organizer: **Ting, Lee Mei-Ling** Harvard Medcal School

Address: Channing Laboratory/BWH Harvard Medical School 181 Longwood Avenue Boston, MA 02115
Phone: 617-525-2732
Fax: 617-731-1541
Email: stmei@channing.harvard.edu

Session Timing: 110 minutes total (Sorry about format):

> Opening Remarks by Chair - 5 or 0 minutes First Speaker - 30 minutes (or 25) Second Speaker - 30 minutes Third Speaker - 30 minutes Discussant - 10 minutes (or none) Floor Discusion - 10 minutes (or 5 or 15)

Session Chair: **Sun, Jianguo** University of Missouri

Address: Dept. of Statistics Univ. of Missouri 321 Math Sci Building Columbia, MO 65211
Phone: 573-882-6667
Fax: 573-445-9725
Email: tsun@stat.missouri.edu

1. Sensitivity Analysis and Non Complance in Randomized Trials

Robins, James M., Harvard School of Public Health

Address: James M. Robins Room 3-823 Harvard School of Public Health 677 Huntington Avenue Boston, MA 02115
Phone: 617-432-0206
Fax: 617-566-7805
Email: robins@hsph.harvard.edu

Abstract: In randomized trials with noncompliance, it is often of interest to try to estimate the effect of treatment had there been complete compliance. However the effect under full compliance is *not identified without strong untestable aasumptions. As a consequence, three approaches are available. First, compute bounds for the effect under full compliance. Second quantify uncertainty in a prior distribution and do a Bayesian analysis. Third conduct a sensitivity analysis. This talk offers a new approach to sensitvity analysis and shows how it can recover both results obtained under the bound approach and the Bayesian approach, while offering greater flexibility and insight. Our appoach derives from a theorem stating that although the odds of becoming noncompliant at each time t conditional on the possibly unobserved "outcome under complete compliance" is not idenified, nonetheless given this "selection bias" odds as a function of t, the distribution of the "outcome under complete compliance" is nonparametrically just identified.Hence in a sensitivity analysis, we vary these "selection bias" odds.*

2. Estimating the Effect of Observed Exposure in Experimental Settings

Goethebeur, Els, University of Ghent, Belgium

Address: Els Goetghebeur University of Ghent Department of Applied Mathematics and Informatics Krijgslaan 281-S9 9000 Ghent BELGIUM

Phone: 0032-(0)9- 264 4811

Fax: 0032-(0)9- 264 4995

Email: Els.goetghebeur@rug.ac.be

Abstract: *In their 1991 paper, Efron and Feldman model the relationship between potential placebo response, and the effect of a possible drug exposure. They allow explicitly for an interaction between placebo response and drug dose effect. In doing so they provide measures of effectiveness in complier subgroups (like the effect of full compliance for the subgroup of patients who were full compliers with their assigned treatment), and efficacy at different potential exposures (like the effect of full compliance exposure over the entire study population). These parameters are estimated under the assumption of some form of exchangeability between compliance with placebo and compliance with treatment. In a simulation, Albert and Demets (1994) demonstrate*

how strong the biases can be in such model, when the assumption of comparable patients in similar compliance subgroups on treatment and placebo are not fulfilled.

In this talk we show that the structural mean model (Robins, 1994, Goetghebeur and Lapp, 1997) can be used to estimate all the parameters of interest in the model of Albert and Demets, provided a baseline covariate can be identified which is associated with placebo response or treatment compliance, but which does not interact with the added treatment effect. These estimators are consistent even for selective noncompliance and noncomparable compliance groups between treatment and placebo. In return, they rely on the parametric description of the 'causal' or structural relationships. Ocne this is correctly specified, the SMM model also allows to find the average relationship between treatment compliance and placebo response. We will move on to discuss sensitivity to this new level of structural assumptions.

3. The Impact of Patient Compliance on Steady State Pharmacokinetics

Wang, Wenping, Janssen Research Foundation

Address:
Phone: 609-730-3453
Fax:
Email: wwang@janwcc1.ssw.jnj.com
Hsuan, Francis, Temple University
Chow, Shein-Chung, Biostatistics and Data Management, Covance

Abstract: Physicians commonly prescribe drug products in a multiple dosage regimen for prolonged therapeutic activity. To study the effect of multiple dosing on drug concentration in blood, researchers often employ a deterministic model with the assumptions that drugs are administered at a fixed dosage, with fixed (usually constant) dosing intervals. In practice, as it is well known in the medical community, patients may not follow such a rigid schedule. Hence, two possible scenarios might occur: patients might not take the prescribed dosage, resulting in variable *dosing size; or they might not adhere to the dosing schedule, resulting in irregular dosing times. This paper intends to lay out a probability framework to model these two types of non-compliance and consequently study their impact on the steady state pharmacokinetics.*

When studying multiple dose pharmacokinetics, the principle of superposition is the key tool. In this paper, the principle of superposition in the presence of non-compliance is formulated generally as a recursive formula. With this formula, we are able to generalize the notion of steady state in multiple dose pharmacokinetics given non-compliance. Using the compliance models and the principle of superposition, important pharmacokinetic parameters are rigorously studied. Factors affecting the steady state trough concentration are characterized through a simulation study. The relationship between the compliance index and the average concentration at steady state is established. This result generalize the classic result about the equality between the single dose area under the curve (AUC) and the multiple dose AUC. Using theophylline (an antiasthma agent) as an example, we demonstrate that non-compliance causes the drug concentration time curve to exhibit an increased fluctuation. The increase in fluctuation due to non-compliance cannot be explained with the use of the classical deterministic multiple dose model. A practical question – the number of doses needed to reach steady state, is answered both asymptotically and in a small sample study.

Discussant: **Zelen, Marvin** Dana Farber Cancer Institute and Harvard School of Public Health

Address: Dana Farber Cancer Institute Department of Biostatistical Sciences Mayer 427 44 Binney Street, Boston, MA 02115-6084

Phone: 617-632-3012

Fax: 617-632-2444

Email: zelen@jimmy.harvard.edu

List of speakers who are nonmembers: None

Chapter 3

ASA Biopharmaceutical (3 + 1 competition)

Program Chair: **Capizzi, Thomas P.**

3.1 asa.biopharm.01

Sponsoring Section/Society: ASA-Biopharm

Session Slot: 10:30-12:20 Tuesday
Estimated Audience Size: 200-250
AudioVisual Request: two overheads

Session Title: Formal Design Considerations for Phase I Clinical Trials

Theme Session: No
Applied Session: no

Session Organizer: **Rosenberger, William F.** University of Maryland

Address: Department of Mathematics, University of Maryland 1000 Hilltop Circle Baltimore, MD 21250
Phone: 410-455-2433
Fax: 410-455-1066
Email: Billr@math.umbc.edu

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 5 minutes First Speaker - 25 minutes Second Speaker - 25 minutes Third Speaker - 25 minutes Discussant - 15 minutes Floor Discusion - 15 minutes

Session Chair: **Rosenberger, William F.** University of Maryland

Address: Department of Mathematics, University of Maryland 1000 Hilltop Circle Baltimore, MD 21250
Phone: 410-455-2433
Fax: 410-455-1066
Email: Billr@math.umbc.edu

1. Designing Studies for Dose-Response
Wong, Weng Ke, University of California, Los Angeles

Address: Department of Biostatistics UCLA School of Public Health Los Angeles, CA 90095-1772
Phone: 310-206-9622
Fax: 310-267-2113
Email: wkwong@sunlab.ph.ucla.edu

Abstract: 'Dose resonse' refers to the regression of a response on a stimulus. We review a number of options for dose-response designs, and compare various designs which may be used in practice. We start with two group designs. Next we introduce basic optimal approximate design theory for simple linear and quadratic regression illustrating different criteria of optimality and their effect on the allocation of the levels of the dose. Then we obtain the efficiencies of these optimal approximate designs and some simple designs which have intuitive appeal (symmetry, equal spacing of treatments, reduced numbers of observations at the highest and lowest doses).

2. Ehrenfest Urn Designs for Quantile Estimation
Flournoy, Nancy, American University

Address: Department of Biostatistics American University Washington D.C. 20016-8050
Phone: 202-885-3127
Fax: 202-885-3155
Email: flournoy@american.edu
Durham, Stephen, University of South Carolina

Abstract: Consider a binary random variable for which the probability of success is a monotone increasing function (called the response function) of some control variable, such as the dose of a drug. Except in the extreme tails, the asymptotic optimal design for estimating a particular quantile of the response function puts all observations at one point which is the unknown target quantile. We describe a randomized version the Ehrenfest urn model that provides a useful appproximation to the asymptotic optimal design of a target quantile. Subjects are sequentially treated according to the color of the ball drawn from the urn; after each treatment a ball is added to the urn whose color depends on the treatment outcome according to an algorithm that centers the treatment disribution around the unknown target quantile. Parametric assumptions regarding the response function are not required. Several estimators of the target quantile are presented. Exact distributional results are obtained. Whem blinding treatments is important, the Randomized Ehrenfest Urn Design may be preferred to the unrandomized procedures such as stochastic approximation, continual reassessment and up and down designs.

3. *Dose Escalation Schemes for Cancer Phase I Trials*
Zacks, Shelemyahu, Binghamton University

Address: Department of Mathematical Sciences
Phone: 607-777-6035
Fax: 607-777-2450
Email: shelly@math.binhamton.edu

Abstract: One of the objectives of Phase I clinical trials is to give patients the largest possible dose they can tolerate in order

to test the reaction of patients to a new drug. Patients entering phase I trials have a certain unknown tolerance distribution for the new drug. The maximum tolerated dose, MTD, is a certain quantile of this tolerance distribution. The problem with clinical trials is that we have to design the doseages under the ethical constraint that the probability of giving a dose greater than the MTD is not greater than a small risk level, alpha. Most of the dose escalation schemes in the literature do not satisfy this constraint. Efficient Bayesian and non-Bayesian escalation schemes satisfying the above constraint are discussed and some results are presented of clinical trials. In particular we discuss modeling which can incorporate information given by covariates.

Discussant: **Ting, Naitee** Pfizer, Inc.

Address: Central Research Division Pfizer, Inc Groton, Ct 06340
Phone: 860-441-4871
Fax: unknown
Email: tingn@pfizer.com
List of speakers who are nonmembers: None

3.2 asa.biopharm.02

Sponsoring Section/Society: ASA-BIOPHARM

Session Slot: 10:30-12:20 Monday
Estimated Audience Size: 200-250
AudioVisual Request: Two Overheads
Theme Session: Yes
Applied Session: Yes

Session Title: Assessment of the U.S. FDA Draft Guidance on Statistical Procedures for Bioequivalence Studies

Session Organizer: **Liu, Jen-pei** National Cheng-kung University

Address: Department of Statistics National Cheng-kung University Tainan, Taiwan 70101
Phone: 886-6-275-7575 ext 53622
Fax: 886-5-234-2469
Email: jpliu@hpstat.ncku.edu.tw

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 5 minutes First Speaker - 25 minutes Second Speaker - 25 minutes Third Speaker - 25 minutes Discussant - 15 minutes Floor Discussion - 15 minutes

Session Chair: **Liu, Jen-pei** National Cheng-kung University

Address: Department of Statistics National Cheng-kung University Tainan, Taiwan 70101
Phone: 886-6-275-7575 ext 53622
Fax: 886-5-234-2469
Email: jpliu@hpstat.ncku.edu.tw

1. Population and Individual Bioequivalence: A View from the Pharmaceutical Science Advisory Committee

Davidian, Marie, North Carolina State University

Address: Department of Statistics Box 8203 North Carolina State Univ. Raleigh NC 27695-8203
Phone: 919-515-1940
Fax: 919-515-7591
Email: davidian@stat.ncsu.edu

Abstract: An ongoing research effort within the Center for Drug Evaluation and Research at the FDA has resulted in new guidelines that will require sponsors to show evidence of individual bioequivalence both of generic and innovator drugs and of innovator products subject to post-approval changes. Throughout the development of the new criteria, members of the FDA working group devoted to this effort made regular presentations to the Advisory Committee for Pharmaceutical Science (formerly *the Generic Drugs Advisory Committee) on matters of rationale and implementation, and the issue was a matter of some debate. I served as the sole statistician on the Committee during this period. I will review the process leading to the new draft guidance, recount the impressions and comments of the members of the Committee, and provide my own perspective on the statistical issues involved.*

2. Evaluation of Individual Bioequivalence: An Industry's Perspective

Chow, Shein-Chung, Covance, Inc.

Address: Executive Director, Biostatistics and Data Management Center for Scientific and Technical Services Covance Clinical and Periapproval Services, Inc. 210 Carnegie Center Princeton, New Jersey 08540-6233
Phone: 609-452-4457
Fax: 609-987-9048
Email: scchow@aol.com

Abstract: Individual bioequivalence as the requirement for approval of generic drug products and post-NDA-approval changes will be addressed with respect to average bioequivalence in terms of scientific merit and true necessity. Focus will be on the aggregate criterion suggested in the FDA draft guidance. Advantages and drawbacks of its corresponding statistical procedures and designs recommended in the draft guidance will be discussed with respect to their statistical properties, difficulty in application, interpretation, and the cost required for any potential gain of additional information. Other procedures based on different criteria for assessment of individual bioequivalence will also be discussed and compared to the FDA methods. Feasibility of the FDA draft guidance will be also discussed.

3. Uncertainty of Estimated Variances in the Determination of Individual Bioequivalence

Endrenyi, Laszlo, University of Toronto

Address: Department of Pharmacology Medical Sciences Building University of Toronto Toronto, Canada M5S, 1A8
Phone: 416-978-2728
Fax: 416-978-6395
Email: l.endrenyi@utoronto.ca

Abstract: Multiperiod crossover trials enable the estimation of intraindividual variances of the test and reference formulations (s2WT and s2WR). However, estimated variance are sensitive to single measurements and, especially, outliers. Moreover, the FDA Guidance suggests that difference (s2WT - s2WR) could be a basis for offering a reward to an improved test product. However, according to the recommended model, a large reward or penalty, could occur just by chance with high probability. This raises questions about the usefulness of the model for its implementation.

Discussant: **Bolognese, James A.** Merck Research Laboratories

Address: Merck Research Labs RY33-404 126 Lincoln Ave. Rahway, NJ 07065
Phone: 732-594-4412
Fax: 732-594-6075
Email: jim_bolognese@merck.com
List of speakers who are nonmembers: 1

3.3 asa.biopharm.03

Sponsoring Section/Society: ASA-BIOPHARM

Session Slot: 4:00- 5:50 Sunday
Estimated Audience Size: 200-250
AudioVisual Request: none
Theme Session: no
Applied Session: yes

Session Title: Immunogenicity and Efficacy Issues in Vaccine Clinical Trials

Session Organizer: **Wiens, Brian** Merck Research Laboratories

Address: Merck Research Laboratories P.O. Box 4, BL3-2 West Point, Pa 19486
Phone: 610-397-2864
Fax: 610-397-2931
Email: brian_wiens@merck.com

Session Organizer: **Lachenbruch, Peter A.** FDA/CBER

Address: FDA/CBER/DBE 1401 Rockville Pike HFM-215 Rockville MD 20852
Phone: 301-827-6055
Fax: 301-827-3529
Email: lachenbruch@a1.cber.fda.gov

Session Timing: 110 minutes total (Sorry about format):

> Opening Remarks by Chair-5 minutes First Speaker - 25 minutes Second Speaker - 25 minutes Third Speaker - 25 minutes Discussant - 15 minutes Floor Discussion - 15 minutes

Session Chair: **Wiens, Brian** Merck Research Laboratories

Address: Merck Research Laboratories P.O. Box 4, BL3-2 West Point, Pa 19486
Phone: 610-397-2864
Fax: 610-397-2931
Email: brian_wiens@merck.com

1. Time-to-Event Models For Vaccine Studies: Old Answers to New Questions

Moulton, Lawrence H., The Johns Hopkins University

Address: Department of International Health Room W5515 (W5602 office) The Johns Hopkins University School of Hygiene and Public Health 615 North Wolfe Street Baltimore, MD 21205
Phone: 410-955-6370
Fax: 410-502-6733
Email: lmoulton@jhsph.edu

Abstract: The routine analysis of vaccine efficacy trials consists of a simple summary of the data into attack rates or risks, and then estimating efficacy as (1-RR)x100%, where RR is the rate or risk ratio for vaccinated versus unvaccinated individuals. For many infectious disease settings, however, and particularly for vaccine studies in infants, these summaries ignore many important analytic features of the data. These include: 1) multiple disease episodes, and the corresponding within-child correlation; 2) seasonality of disease; 3) staggered enrollment; 4) other time-varying covariates, including age; 5) waning of vaccine efficacy. Such features may be accounted for using various implementations of the Cox proportional hazards model. Today's presentation illustrates how these techniques may be used in vaccine efficacy trials. Data from a rotavirus vaccine study are used to estimate efficacy in the face of multiple events, and to estimate the degree of immunity afforded by natural infection. Hib vaccine efficacy and effectiveness studies are combined to estimate the impact of indirect effects of the vaccine in reducing incidence in a closed population.

2. A Theroretical Framework for the Relationship of Correlates of Protection, Population Immunogenicity, and Vaccine Efficacy

Kohberger, Robert C., Wyeth-Lederle Vaccines and Pedatrics

Address: N. Middletown Rd. Pearl River, NY 10965
Phone: 914-732-4249
Fax: 914-732-5517
Email: Robert_Kohberger@internetmail.pr.cyanamid.com
Xie, Fang, Wyeth-Lederle Vaccine and Pediatrics

Abstract: Typical vaccine clinical trials study immunogencity or efficacy and sometimes both. In an immunogenicity trial, subjects are treated with an experimental vaccine and then antibody levels are measured at various times. There are three major types of outcome measurements of these trials: 1) pre vaccination and post vaccination titers are taken and the % seroconverters are used as the outcome measurement. 2) post vaccination titers are *taken and the % above a certain level are used as the outcome measurement. 3) post vaccination titers themselves are used as the outcome measurement. This discussion will focus on the situation where the post vaccination titer, either as an absolute value or % above a certain level, is the outcome measurement. In an efficacy trial, subjects are treated with an experimental vaccine and disease rates compared between the treated group and a control group. This type of trial measures vaccine efficacy. The immunogenicity trial and the efficacy trial are linked with the concept of a correlate of protection. It is assumed that if populations have equivalent antibody titer distributions, they will have equivalent efficacy. This is true only if there is a linkage between efficacy and immunogenicity through protective correlates. For immunogenicity trials where percentages are the outcome measure then equivalency is determined not by distributions but by a single value. An equivalency trial has an implied concept of a difference in titer distributions that is not clinically relevant. In order to determine clinical relevancy, the link between efficacy and immunogenicity must be determined. This paper will develop an analytic framework that will combine correlates of protection, population antibody titer distribution and vaccine efficacy. Examples of typical correlate models and immunogenicity trials will be provided to demonstrate the effect of changes in titer distributions on vaccine efficacy.*

3. Issues in Predicting the Duration of Vaccine Induced Protection

Bohidar, Norman R., Merck Research Laboratories

Address: Merck Research Laboratories P.O. Box 4, BL3-2 West Point, PA 19486
Phone: 610-397-2981
Fax: 610-397-2931
Email: Norman_Bohidar@merck.com
Pigeon, Joseph G., Villanova University
Wiens, Brian, Merck Research Laboratories

Abstract: When new vaccine products are marketed, they are generally supported by clinical studies lasting only several years. However, the duration of protection from disease is often expected to be considerably longer than the time spans of the clinical studies. Thus there are both methodological and conceptual problems in trying to estimate long term protection based on data from short term clinical studies. We discuss the utility of predicting the duration of protection based on extrapolating observed antibody data. Several methods that have appeared in the clinical and statistical literature are reviewed and some alternative strategies are proposed. A sample data set is used to illustrate these competing models. Finally, we discuss the advantages and disadvantages of the various methods and make some recommendations about predicting the duration of vaccine-induced protection from short term antibody data.

Discussant: **Lachenbruch, Peter A.** FDA-CBER

Address: FDA/CBER/DBE 1401 Rockville Pike HFM-215 Rockville MD 20852
Phone: 301-827-6055
Fax: 301-827-3529
Email: lachenbruch@a1.cber.fda.gov
List of speakers who are nonmembers: None

3.4 asa.biopharm.04

Sponsoring Section/Society: ASA-Biopharm

Session Slot: 8:30-10:20 Thursday
Estimated Audience Size: 150-200
AudioVisual Request: Two Overheads

Session Title: Permutation Tests in Clinical Trials

Theme Session: No
Applied Session: Yes

Session Organizer: **Berger, Vance**   Food and Drug Administration

Address: CBER Food and Drug Administration 5600 Fishers Lane Rockville MD 20857
Phone: 301-827-3977
Fax: 301-827-3529
Email: bergerv@a1.cber.fda.gov

Session Timing: 110 minutes total (Sorry about format):

> Opening Remarks by Chair - 5 minutes First Speaker - 25 minutes Second Speaker - 25 minutes Third Speaker - 25 minutes Fourth Speaker - 25 minutes Discussant - None Floor Discusion - 5 minutes

Session Chair: **Berger, Vance**   Food and Drug Administration

Address: CBER Food and Drug Administration 5600 Fishers Lane Rockville MD 20857
Phone: 301-827-3977
Fax: 301-827-3529
Email: bergerv@a1.cber.fda.gov

1. Issues on Randomization Tests: Applications in Analysis of Ischemic Stoke Lesion Volume in the NINDS t-PA Strok Trial

Lu, Mei, Henry Ford Health System

Address: Department of Biostatistics and Research Epidemiology Henry Ford Health System 3E 1 Ford Place, Detroit, MI 48202
Phone: 313-874-6413
Fax: 313-874-6730
Email: mlu1@biostat.hfh.edu
Tilley, Barbara, Henry Ford Health System
Li, Shuhui, Henry Ford Health System

Abstract: The NINDS Stroke Trial demonstrated the effectiveness of tissue-type plasminogen activators (t-PA) given within three hours of stroke onset for the treatment of acute ischemic *stroke. This effectiveness was defined as clinical recovery measured at three months. Assessment of the impact on stroke lesion volumes measured at three months is desirable. We initially proposed an analysis for the effect of t-PA on stroke lesion volume using randomization tests, given the highly skewed data and the distribution-free assumption for randomization tests. However, we encountered difficulties in analyzing th estroke lesion volume. Results from randomization tests are dependent on test statistics calculated from observed data. In this paper, we discuss the aspects of randomization tests and their limitations, illustrate our findings in analyzing the stroke lesion volumes for the effect of t-PA, and compare randomization tests with nonparametric statistics. To study the effect of t-PA on the stroke lesion volumes, we finally proposed a parametric analysis using a generalized estimate equations (GEE) approach on the transformed data and validated the results using randomization tests with a relevant test statistic. This experience with the NINDS t-PA trial data emphasizes that randomization tests are useful for clinical trials but have limitations. The choice of test statistics influences randomization tests, and the choice of test statistics should be appropriate to the given data and hypotheses. Randomization tests are equivalent if test statistics are equivalent.*

Key Words: Randomization Test; Permutation Test; Exact Test; Nonparametric Test; GEE; Robustness; Statistical Validation.

2. A New Look at Rank Tests in Ordered 2 x k Contingency Tables

Permutt, Thomas, Food and Drug Administration

Address: FDA/CDER 5600 Fishers Lane (HFD-170) Rockville MD 20857

Phone: 301-443-3741

Fax: 301-443-7068

Email: PermuttT@cder.fda.gov

Berger, Vance, Food and Drug Administration

Abstract: *The P-P plot allows properties of rank tests in trials with two groups and ordinal outcomes to be visualized. We consider several tests that have been proposed in this application (Smirnov, Wilcoxon, and maximum chi-square). We compare them to the van der Waerden test, well known but not in this setting, and a new, cumulative chi test. We conclude that rank tests are a viable alternative to a priori assignment of parametric scores for outcomes. Both Smirnov-like tests and Wilcoxon-like tests may be useful. However, both the Smirnov and Wilcoxon tests are bad examples of their types in that they give equal weight to some deviations that are highly improbable under the null hypothesis and to others that are much less improbable. The maximum chi-square test is more powerful than the Smirnov test, and the van der Waerden and cumulative chi tests are more powerful than the Wilcoxon test, over a wide region of the parameter space.*

3. A Comparison of One-Sided Methods to Identify Significant Individual Outcomes in a Multiple Outcome Setting

Troendle, James, National Institute of Child Health and Development

Address: National Institutes of Health Bld. 6100, Rm. 7B13 Bethesda, MD 20892

Phone: 301-496-6811

Fax: 301-402-2084

Email: jft@helix.nih.gov

Legler, Julie L., National Cancer Intstitute

Abstract: *We compare two approaches to the identification of individual significant outcomes when a comarison of two groups involves multiple outcome variables. The approaches are all designed to control the familywise error rate in the strong sense. The first approach is initially to use a global test of the overall hypothesis that the groups are equivalent for all variables, followed by an application of the closed testing algorithm. The global tests considered here are ordinary and generalized least squares, an appoximation to the likelihood ratio test, and an approximation to* *the most powerful similar test for a simple normal alternative. The second approach is that of stepwise testing, which tests the univariate hypotheses in a specific order with appropriate adjustment to the univariate p-values for multiplicity. The stepwise tests considered are those of Holm and Hochberg, along with step-down and step-up permutation tests. The permutations allow the tests to incorporate the correlation between outcomes without any restrictive assumptions.*

We illustrate the tests with two examples of birth outcomes: a comparison of cocaine exposed newborns to control newborns on neurobehavioral and physical growth variables, and, in a separate study, a comparison of babies born to diabeticmothers and babies born to non-diabetic mothers on minor malformations. Simulations are used to provide a basis for chosing among the procedures considered.

4. ***Exact Statistical Inference in Familial Disease Clusters***

Zelterman, Dan, Yale University

Address: Division of Biostatistics PO Box 208034 Yale University New Haven, CT 06520
Phone: 203-737-1768
Fax: 203-785-4116
Email: Daniel.Zelterman@Yale.edu
Yu, Chang, Yale University
Claus, Elizabeth B., Yale University

Abstract: In many epidemiologic studies the first indication of an environmental or genetic contribution to the disease is the way in which the diseased cases cluster within the same family units. We assume that all individuals are exchangeable, except for their disease status. This assumption of exchangeability is used to test the initial hypothesis of no familial link with the disease. Conditional on the distribution of the sizes of the various familial units, we obtain the exact probability of observing a given set of disease cluster frequencies. The expected frequencies and other moments are given. Three numerical examples demonstrate these *methods. We describe an algorithm for obtaining exact statistical inference by enumerating all possible outcomes consistent with the numbers and sizes of the family units. A parametric model is given to identify useful risk factors for the disease outcome.*

Keywords: exact inference; genetics; statistical computing; multivariate discrete distributions.

List of speakers who are nonmembers: None

Chapter 4

ASA Business and Economic Statistics (4)

Program Chair: **Ericsson, Neil R.**

4.1 asa.business.01

Sponsoring Section/Society: ASA-Business and Economic Statistics

Session Slot: 2:00- 3:50 Sunday
Estimated Audience Size: 100 dws
AudioVisual Request: none dws

Session Title: Multiple Imputation of Income and Assets in Economic Surveys: Application and Assessment

Theme Session: Yes/No
Applied Session: Yes/No

Session Organizer: **Paulin, Geoffrey D.** Bureau of Labor Statistics

Address: BLS-DCES 2 Mass. Ave. NE (PSB 3985) Washington DC 20212
Phone: 202-606-6870 x200
Fax: 202-606-7006

Email: paulin_g@bls.gov

Session Timing: 110 minutes total (Sorry about format):

110 minutes total Opening Remarks by Chair - 5 or 0 minutes First Speaker - 30 minutes (or 25) Second Speaker - 30 minutes Third Speaker - 30 minutes Discussant 1 - 10 minutes (or none) Discussant 2 - 10 minutes (or none) Floor Discusion - 10 minutes (or 5 or 15)

Session Chair: **Rubin, Donald B.** Harvard University

Address: Cambridge, MA 02138
Phone:
Fax:
Email:

1. Multiple Imputation of Income in the Consumer Expenditure Survey: Evaluation of Statistical Inferences

Raghunathan, Trivellore E., University of Michigan

Address:
Phone:
Fax:
Email: teraghu@umich.edu
Paulin, Geoffrey D., Bureau of Labor Statistics

Abstract: *Non-response is a problem common to many surveys. For example, many respondents fail to report incomes for some or all working members of their families in the U.S. Consumer Expenditure Interview Survey. Because these data are so important to economic and other analyses, a complete set of data is desirable. The U.S. Bureau of Labor Statistics and U.S. Bureau of the Census have conducted investigations that use model-based, multiple imputation methods as a viable way of obtaining valid inferences when the data are subject to nonresponse. In the course of this research several approaches have been examined for* *both theoretical and practical reasons (including feasibility of implementation). This paper first examines four approaches: "Partial" Bayesian (with and without total expenditures as a predictive variable) and "Fully" Bayesian (with and without total expenditures as predictor). Although the "Fully" Bayesian model with total expenditures as predictors is found to be the best model of those tested from a statistical standpoint, the use of expenditures as predictors in the imputation model, at least in theory, may have implications for econometric analyses using these data. The second part of this paper examines the implications of not including expenditures as predictors in the imputation model by analyzing inferences based on certain econometric models.*

2. Multiple Imputation in the Survey of Consumer Finances: Experience from the 1989-1998 Surveys

Kennickell, Arthur, Federal Reserve Board

Address: 20th and C Streets NW, Wash. DC 20551
Phone: 202-452-2247
Fax:
Email: m1abk00@FRB.GOV;

Abstract: *The SCF collects information on a broad range of financial variables, many of which are highly skewed in the population. Given the sensitive nature of some of the survey questions, item nonresponse is a serious concern.*

Beginning in 1989, the Survey of Consumer Finances (SCF) has used multiple imputation (MI) to compensate for missing data. This development arose from a compelling need for logically consistent procedures that would not mask the level of uncertainty in imputation. Overall, the SCF experience with MI has been very successful. This paper discusses the implementation of MI for the SCF, and the reception of multiply-imputed data by users.

Discussant: **Bradley, Ralph** Bureau of Labor Statistics

Address: 2 Mass. Ave. NE Wash, DC 20212

Phone: 202-606-6573
Fax:
Email:

Discussant: **Folsom, Jr., Ralph E. (not confirmed)** Research Triangle Institute

Address: Research Triangle Institute PO Box 12194 RTP, NC 27709-2194
Phone:
Fax:
Email:
List of speakers who are nonmembers: None

4.2 asa.business.02

Sponsoring Section/Society: ASA Business & Economics, ASA Health Policy Statistics, and ASA Committee on National & International Security

Session Slot: 2:00- 3:50 Tuesday
Estimated Audience Size: 75-125
AudioVisual Request: any?

Session Title: Assessment of Drug Interdiction Activities

Theme Session: Yes
Applied Session: Yes

Session Organizer: **Fries, Arthur** Institute for Defense Analyses

Address: IDA, 1801 N. Beauregard St., Alexandria, VA 22311
Phone: (703) 845-2364
Fax: (703) 845-2274
Email: afries@ida.org

Session Timing: 110 minutes total (Sorry about format):

4 speakers Opening Remarks by Chair - 3 minutes First Speaker - 25 minutes Second Speaker - 25 minutes Third Speaker - 25 minutes Fourth Speaker - 25 minutes Floor Discussion - 7 minutes

Session Chair: **Spruill, Nancy**   Office of the Secretary of Defense

Address: Deputy Director, Acquisition Resources Room 1E474 Pentagon, Washington, DC 20310
Phone: (703) 693-2982
Fax: (703) 695-2086
Email: spruilnl@acq.osd.mil

1. ***NRC Study–Information and Research for Drug Abuse Policy***
Thomas, Yonette,   National Research Council

Address: Committee on Law and Justice HA156L National Research Council 2101 Constitution Ave. NW Washington DC 20418
Phone: (202) 334-1702
Fax: (202) 334-3829
Email: ythomas@nas.edu

Abstract: The National Research Council has begun a 30-month study entitled: "Data and Research for Policy on Illegal Drugs," sponsored by the Office of National Drug Control Policy. The study will focus on resolving conflicts among existing cost-effectiveness studies that cut across interdiction, enforcement, treatment, and on specifying the data and research required for improved policy studies to assess the effectiveness of various drug control strategies. The study is expected to identify information that can stimulate the development of both new data collection systems and advanced models that can respond to changing policy analysis needs. This talk will describe the study committee's task and provide an update on the study efforts to date.

2. ***The Need for More Drug War Statistics***
Fries, Arthur,   Institute for Defense Analyses

Address: IDA, 1801 N. Beauregard St., Alexandria, VA 22311
Phone: (703) 845-2364
Fax: (703) 845-2274
Email: afries@ida.org

Charles, Robert, House Subcommittee on National Security, International Affairs and Criminal Justice

Abstract: We will address: (1) the rising importance of interdiction in the context of US Drug Policy, as increasingly framed by Congress and the Executive (in historical context); (2) the promise of simultaneously marrying stronger intelligence with a more robust maritime, border and source country interdiction effort; and (3) the need for accurate data collection in order to both make interdiction more effective and assure greater overall accountability.

3. Models Versus Data

Rivolo, Rex, Institute for Defense Analyses

Address: IDA, 1801 N. Beauregard St., Alexandria, VA 22311
Phone: (703) 845-6760
Fax: (703) 845-2274
Email: arivolo@ida.org

Abstract: We provide a historical outline of research in the field of illegal drug use, treatment, and prevention, with a focus on recent published reports that have played major roles in the development of national and international drug interdiction policies. Reported results and conclusions from this body of research are shown to suffer from fundamental analysis flaws and to be inconsistent with empirical evidence. In particular, specifications of predetermined models appear to dominate over straightforward data presentations and summaries.

4. Variation in the Optimal Mix of Drug Control Interventions over the Course of a Drug Epidemic

Caulkins, Jonathan, Carnegie Mellon University

Address: Carnegie Mellon University, The Heinz School of Public Policy and Management, Pittsburgh PA 15213

Phone: (412) 268-3840
Fax: (412) 268-7036
Email: caulkins+@andrew.cmu.edu
Feichtinger, Gustav, Vienna University of Technology
Tragler, Gernot, Vienna University of Technology
Behrens, Doris, Vienna University of Technology
Haunschmied, Josef, Vienna University of Technology

Abstract: There is considerable interest in what the optimal mix of drug control interventions is, but to date most work has not explicitly acknowledged the epidemic character of waves of drug use or the possibility that the optimal mix of interventions should change over time. This paper applies optimal control theory to some very simple models of drug epidemics to produce insights into this important policy problem.

List of speakers who are nonmembers: Thomas, Rivolo, Caulkins

4.3 asa.business.03

Sponsoring Section/Society: ASA B&E

Session Slot: 2:00- 3:50 Wednesday
Estimated Audience Size: xx-xxx
AudioVisual Request: xxx

Session Title: Empirical Auction Models

Theme Session: Yes/No
Applied Session: Yes/No

Session Organizer: **Marshall, Robert C.** Pennsylvania State University

Address:
Phone:
Fax:
Email:

Session Timing: 110 minutes total (Sorry about format):

110 minutes total Opening Remarks by Chair - 0 or 5 minutes First Speaker - 30 minutes (or 25) Second Speaker - 30 minutes Third Speaker - 30 minutes Discussant - 10 minutes (or none) Floor Discussion - 10 minutes (or 5 or 15)

Session Chair: **Marshall, Robert C.** Pennsylvania State University

Address: Department of Economics, 613 Kern Graduate Building, Penn State University, University Park, PA 16802
Fax:
Email: rcm10@psu.edu

1. Econometrics of the First Price Auction with Asymmetric Bidders

Bajari, Pat, Havard University

Address: 200 Littauer Center, Department of economics, Harvard University, Cambridge, Ma 02138
Phone: 617-495-8040
Fax: 617-495-8570
Email: pbajari@arrow.fas.harvard.edu

Abstract: This research develops an approach to econometric modeling of the first price sealed bid auction with independent private values. Previous econometric work imposes the assumption that the distribution of bidder valuations is identical for all bidders. This research allows bidders to draw valuations from potentially different distributions. Thus heterogeneity of bidders is allowed. The approach to inference in this research is Bayesian. Difficulties in classical inference, such as nonstandard asymptotic theory are therefore circumvented. Existence and uniqueness of equilibrium for the theoretical model is established. The general econometric approach is then applied to the specific problem of asphalt firms bidding for government contracts in the vicinity of Minneapolis, Minnesota. This research finds that both location *and differing levels of efficiency across firms are important factors in determining the price level. Estimates for firm profitability are produced. The structural model is compared to a linear specification. Non-nested hypothesis testing strongly favors the structural model. A test for bid rigging in the market is proposed. The bid rigging test differs from previous approaches in that an equilibrium model of competition and an equilibrium model of collusion are directly compared. The results strongly favor the hypothesis of competition. Finally, this research compares expected price levels under both a first price and second price auction.*

2. The Impact of Synergies on Bidding in the Georgia School Milk Market

Marshall, Robert C., Pennsylvania State University

Address: Department of Economics, 613 Kern Graduate Building, Penn State University, University Park, PA 16802

Phone:

Fax:

Email: rcm10@psu.edu

Raiff, Matt E., Pennsylvania State University

Abstract: Each summer milk processors around the country participate in sealed bid procurements for the right to provide public schools with milk throughout the subsequent academic year. School district contracts are an important part of a complicated vehicle routing problem that milk processors solve on an ongoing basis. There are allegedly substantial cost savings for a milk processor from servicing a district that is directly adjacent to one they already service. In this paper, following the work of Krishna and Rosenthal (1996), we construct a procurement model allowing for cost synergies. The equilibrium bid function, which reproduces the single object first price procurement bid when synergies are not present, maps directly into an empirical specification. Using data from a time period when bidders were allegedly acting non-cooperatively we find significant and substantial support for the presence of synergies in the bidding. This implies that models of

collusive bidder behavior must take into account the presence of synergies in bidder valuations.

3. Identification and Estimation of a Class of Game Theoretic Models

Florens, Jean-Pierre, Univ. des Sciences Sociales, Toulouse

Address:
Phone:
Fax:
Email:
Protopopescu, C., Univ. des Sciences Sociales, Toulouse
Richard, Jean-Francois, University of Pittsburgh

Abstract: A large class of game theoretic models is represented by the following statistical model: $\xi_i \sim F$ i.i.d. and $x_i = \varphi(\xi_i, F)$. The $x_i's$ are observable but the $\xi_i's$ are latent variables. The function φ is known and represents the strategy of players. The strategical component is expressed by the dependence of ξ upon F. More sophisticated versions of the model are obtained by introducing exogenous variables in the distribution of ξ_i and/or in the function φ. The paper considers parametric and nonparametric estimation of the distribution function F. Issue of particular interest are the identification properties of the model and, in the nonparametric case, the speed of convergence of the estimator of F. The latter depends upon the properties of the game and leads to define a concept of (statistical) order of game.

List of speakers who are nonmembers: None???

4.4 asa.business.04

Japanese Statistical Society

Session Slot: 10:30-12:20 Tuesday
Estimated Audience Size: ???

AudioVisual Request: Not sure

Session Title: Modeling and Model Selection: Theory and Applications

.

Theme Session: Yes
Applied Session: Yes

Session Organizer: **Kitagawa, Genshiro** The Institute of Statistical Mathematics

Address: The Institute of Statistical Mathematics 4-6-7 Minami-Azabu Minato-ku, Tokyo 106 Japan
Phone: +81-3-5421-8746
Fax: +81-3-3446-1695
Email: kitagawa@ism.ac.jp

Session Timing: 110 minutes total (Sorry about format):

> 110 minutes total....arranged as (please edit) Opening Remarks by Chair - 5 or 0 minutes First Speaker - 30 minutes Second Speaker - 30 minutes Third Speaker - 30 minutes Discussant - 10 minutes Floor Discussion - 10 minutes

Session Chair: **Monsell, Brian C.** U. S. Bureau of the Census

Address: Statistical Research Division, Room 3000-4, Washington, DC 20233
Phone: 301-457-2985
Fax: 301-457-2299
Email: bmonsell@census.gov

1. Statistical Model Evaluation and Information Criteria
Konishi, Sadanori, Kyushu University

Address: Graduate School of Mathematics, Kyushu University, 6-10-1 Hakozaki, Higashi-Ku, Fukuoka 812-81, Japan
Phone: 81-92-642-2766

Fax: 81-92-642-2779
Email: konishi@math.kyushu-u.ac.jp

Abstract: *The problem of evaluating the goodness of statistical models is fundamental and of importance in various fields of statistics. Akaike's (1973, 1974) information criterion, known as AIC, provides a useful tool for constructing statistical models. AIC was proposed as an estimate of the expected Kullback-Leibler discrepancy between the true model generating the data and a model estimated by maximum likelihood. With the development of various modeling techniques, the construction of criteria which enable us to evaluate various types of statistical models has been required. We present a unified information-theoretic approach to statistical model evaluation and selection problems, and introduce generalized information criteria, relaxing the assumptions imposed on AIC. The criteria are applied to the evaluation of the various types of statistical models based on robust, maximum penalized likelihood, Bayes procedures, etc. We discuss the use of the bootstrap methods in model evaluation, and also consider the problem of choosing the smoothing parameter for spline smoothing in generalized linear models from an information-theoretic point of view.*

2. Modeling of Aftershocks and Change-Point Analysis

Ogata, Yosihiko, The Institute of Statistical Mathematics

Address: 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106-8569 Japan
Phone: +81-3-5421-8744
Fax: +81-3-3446-1695
Email: ogata@ism.ac.jp

Abstract: *In many cases an aftershock sequence is more complex than a nonstationary Poisson process with the intensity rate of an inverse power decay. A history dependent point process called the ETAS model proposed to allow for the effects of varying background seismicity is better fitted to most aftershock sequences, including non-volcanic type swarms. We summarize the* *models, and then focus on the examination of change-point in the aftershock activity which leads to relative quiescence: a significant seismicity lowering from the occurrence rate expected by the estimated ETAS model. Significance of the change-point is discussed from the viewpoint of a model selection which is actually nonstandard. In many cases, aftershocks decrease without any change-point but, in other cases, we find relative quiescence, which is very often followed by another large event in the neighborhood. This research has been initiated by the hope that the relative quiescence can be helpful in predicting whether or not a large event follows shortly in the neighborhood of the aftershock area.*

3. Statistical Models of Molecular Evolution and Comparison

Kishino, Hirohisa, University of Tokyo

Address: Department of Advanced Social and International Studies, University of Tokyo, 3-8-1 Komaba, Meguro-ku, Tokyo 153-8902 Japan
Phone: +81-3-5454-6472
Fax: +81-3-5454-4339
Email: kishino@waka.c.u-tokyo.ac.jp
Thorne, Jeffrey L., Department of Statistics, North Carolina State University
Bar-Hen, Avner, Campus International de Baillarguer

Abstract: *DNA and protein sequence data yield more reliable estimates of evolutionary relationships than do morphological data sets. Maximum likelihood estimation of phylogenetic topology can be accomplished by the adoption of an explicit Markov model for the process of sequence evolution. With maximum likelihood approaches, a balance between the bias and variance of topology estimates can be achieved.*

Although a constant rate of sequence evolution need not be assumed for phylogenetic inference, this assumption of a perfect "molecular clock" has traditionally been made when attempting to infer important historical dates. By employing a hierarchical model for change of evolutionary rates, dates can sometimes be

well estimated even in the absence of the constant rate assumption.

Statistical comparisons between topologies can be regarded as comparisons between separate families of hypotheses. We estimate the variances and covariances of the log likelihood ratios from the sample variances and covariances between the log likelihood ratios of sites. A parametric and a nonparametric bootstrap approach were developed to evaluate the precision of estimated topologies. Difficulty in the interpretation of bootstrap techniques applied to phylogeny reconstruction can be overcome via a hierarchical model with a Yule process prior for lineage divergence.

Discussant: **Findley, David** Bureau of the Census

Address: Statistical Research Division, Room 3000-4, Washington, DC 20233-9100
Phone: 301-457-4983
Fax: 301-457-2299
Email: dfindley@census.gov
List of speakers who are nonmembers: None

Chapter 5

ASA Epidemiology (2)

Program Chair: **D'Agostino, Jr., Ralph B.**

5.1 asa.epid.01

ASA-Epidemiology

Session Slot: Monday, 10:30 - 12:20
Estimated Audience Size: 200
AudioVisual Request: Not sure

Session Title: Issues Arising in the Use of Epidemiologic Studies as Evidence in Legal and Regulatory Decision-Making

Theme Session: YES
Applied Session: Yes

Session Organizer: **D'Agostino Jr., Ralph** Wake Forest University School of Medicine

Address: Dept of Public Health Science/ Biostatistics Medical Center Blvd. Winston-Salem, NC 27157-1063
Phone: (910) 716-9410
Fax: (910) 716-5425
Email: rdagosti@rc.phs.bgsm.edu

Session Organizer: **Gastwirth, Joseph L.** George Washington University

Address: Department of Statistics 2201 G St. NW Washington, D.C. 20052
Phone: (202) 994-6548
Email: jlgast@research.circ.gwu.edu

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 5 or 0 minutes First Speaker - 30 minutes (or 25) Second Speaker - 30 minutes Third Speaker - 30 minutes Discussant - 10 minutes (or none) Floor Discusion - 10 minutes (or 5 or 15)

Session Chair: **Reynolds, Gladys H.** Centers for Disease Control and Prevention

Address: 1600 Clifton Rd., NE Atlanta, GA 30333
Phone: (404)639-7224
Fax: (404)639-7039
Email: GHR1@CDC.GOV

1. ***Casual Inference in Legal Context***
Loue, Sana, Case Western University

Address: Dept of Epidemiology and Biostatistics Ohio

Abstract: This presentation first discusses the differing purposes of law and epidemiology. It next focuses on the elements required to establish causation in a tort case and briefly summarizes the various legal mechanisms that can be sued to establish or refute causation. These legal criteria are compared to epidemiological principles and methods used in the process of causal inference. The presentation concludes with a brief discussion of the role of expert witnesses, such as epidemiologists and biostatiticians, in litigation.

2. What Does It Mean To Estimate The Causal Effects Of Smoking

Rubin, Donald B., Harvard University

Address: Dept of Statistics One Oxford St. Cambridge, MA 02138
Phone: (617)495-5498
Fax: (617) 496-8057
Email: rubin@stat.harvard.edu

Abstract: *Most believe that smoking is a major "cause" of certain disease such as lung cancer. Current litigation, however, concerns the health-care costs "caused by smoking", in particular, those costs caused by the (alleged) misconduct of the tobacco industry starting roughly a quarter century ago. What does it mean to try to estimate these effects? And what does it mean to try to estimate the causal effects on health-care costs of "the existence of smoking" ignoring any issues of misconduct? In what ways are these different questions?*

3. Some Suggestions for Reconciling the Values of Statistical Science With the Goals and Needs of the Legal and Regulatory Processes

Gastwirth, Joseph L., George Washington University

Address: Department of Statistics 2201 G St. NW Washington, D.C. 20052
Phone: (202) 994-6548
Email: jlgast@research.circ.gwu.edu

Abstract: *In principle, science is an ongoing search for the truth. Hence, scientific knowledge is contingent in the sense that any theory is subject to falsification if and when predictions logically derived from it are not borne out in an appropriate experiment. The legal system, on the other hand, is a process for resolving disputes in a reasonably prompt, fair and socially acceptable manner. Epidemiologic studies are used in product liability* *cases which are filed after a harm has occurred. Most states have statues of limitations requiring plaintiffs to file a claim within a pre-set time after the harm occurred or after it was discovered by them. Regardless of the state of scientific knowledge at the end of the limitations period, if plaintiffs have not filed suit, they are precluded from receiving compensation no matter how strongly subsequent studies demonstrate a causal relationship between the product in question and the harm. Similarly, a manufacturer's liability, especially on their duty to warn users of a potential hazard, depends on the state of knowledge available at the time it made and sold the product. Unlike the open-ended nature of scientific investigation, courts need to reach a decision based on the evidence available at the time of the trial. Data from two cases illustrates the point that a change in the view point of the scientific community due to the developments subsequent to a trial does not necessarily imply that the original legal decisions were wrong. Similarly, government officials must decide whether to recall a product or issue a warning based on the available information, knowing that the public may be exposed to a potential harm during the time it takes for new studies to be conducted. After a brief review of the studies that lead to the warning campaign that eventually reduced the incidence of Reye's syndrome by over 80%, suggestions for balancing the trade-off between maximal scientific knowledge and the need to protect the public will be given.*

4. The Impact of Daubert v. Merrell Dow Pharmaceuticals on the Admissibility of Statistical-Influential Evidence of Causation.

Green, Michael, Univeristy of Iowa

Address: Law School Iowa City, Iowa

Abstract: The legal system has long relied on expert witnesses to provide opinion about causal phenomena that are not subject to other readily available proof. One of the most significant areas where this occurs is when a person with disease claims that is was caused by a toxic agent. Even before the revolutionary asbestos litigation began in the mid- 1970's, experts and statistical evidence

had been employed in cases involving exposure to the svine-flu vaccine and to radiation.

As toxic substance litigation grew, increasing concern was raised about the quality of expert testimony. The adversary system tends to pressure even honest experts to state their conclusions more surely to omit qualifications that might enhance the validity of their opinion, and to draw tenuous inferences. Physicians were sometimes permitted to testify about causation based on clinical experience, which in some cases was no more than a part hoc ergo propter hoc rationale.

The United States Supreme Court's decision in Daubert v. Merrell Dow Pharmaceuticals in 1993 imposed significant limits on the admission of expert scientific opinion. The methodology and the reasoning of the expert are to be examined by the court for scientific validity. Unless the trial judge decides that the expert's opinion passes this threshold test, the expert's testimony may not be admitted in court.

The precise scope and impact of the Court's opinion on the admissibility of expert testimony about causation in toxic substance cases remains unclear and controversial. This paper will examine the ambiguity created by the Daubert case and the way in which lower courts are resolving that ambiguity.

List of speakers who are nonmembers: Michael Green

5.2 asa.epid.02

ASA Competition: Section on Epidemiology

Session Slot: 8:30-10:20 Wednesday
Estimated Audience Size: xx-xxx
AudioVisual Request: xxx

Session Title: Estimating the Magnitude of Transmissible Spongiform Encephalopathy Epidemics in the United Kingdom: Bovine Spongiform Encephalopathy (BSE) and the Probable Link to the New Variant of Creutzfeldt-Jakob Disease (vCJD) in Humans

Theme Session: Yes
Applied Session: Yes

Estimation techniques for the magnitude of the BSE epidemic in Great Britain will be presented. The Age-at-onset approach estimates an upper bound to the epidemic with few parametric assumptions. The backcalculation approach simultaneously estimates necessary parameters and the size of the epidemic. The third presentation will address the problem of estimating the magnitude of the vCJD epidemic in humans given that the incubation distribution is long, variable and unknown.

Session Organizer: **MaWhinney, Samantha** University of Colorado Health Science Center

Address: Department of Biometrics, Campus Box B119 University of Colorado Health Science Center 4200 East Ninth Ave Denver CO 80262
Phone: 303 315 7509
Fax: 303 315 3183
Email: sam@greeneggs.uchsc.edu

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 5 or 0 minutes First Speaker - 30 minutes (or 25) Second Speaker - 30 minutes Third Speaker - 30 minutes Discussant - 10 minutes (or none) Floor Discussion - 10 minutes (or 5 or 15)

1. Estimating an Upper Bound on the BSE Epidemic in British Cattle

Samantha, MaWhinney, University of Colorado Health Science Center

Address: Department of Biometrics, Campus Box B119 University of Colorado Health Science Center 4200 East Ninth Ave Denver CO 80262
Phone: 303 315 7509
Fax: 303 315 3183
Email: sam@greeneggs.uchsc.edu

Abstract: *The Bovine Spongiform Encephalopathy (BSE) epidemic in British cattle has generated public health concerns and has significantly impacted the British agricultural industry. Backcalculation methods have been used to estimate infection incidence requiring distributional forms for incubation period and age at infection, allowing for maternal and horizontal transmission. Assuming all infections occur early in life, an upper bound is obtained for the number of infections without separately parametrising these distributions. This method relies on a generalised linear model estimated using an EM algorithm to account for the truncated data. As the epidemic matures, the average age at infection decreases and stabilises (a function of exposure and the feed ban imposed in July 1988) and the upper bound approaches the actual number of infections. The primary analysis was restricted to the birth cohorts of cattle from 1984 to 1993. The bound for the total number of infections was estimated to be 1,061,400, which is consistent with earlier backcalculation results for all cohorts of 903,000 (840,000-1,250,000). The method is shown to yield robust estimates for the magnitude of the epidemic, assuming all infections occur early in life, without requiring the estimation of epidemiological parameters.*

2. Statistical Analysis of the British BSE Epidemic: Past, Present and Future

Donnelly, Christl, University of Oxford

Address: Wellcome Trust Centre for the Epidemiology of Infectious Disease Department of Zoology University of Oxford South Parks Road Oxford OX1 3PS

Phone: 011 44 1865 281 244

Fax: 011 44 1865 281 241

Email: christl.donnelly@zoology.oxford.ac.uk

Abstract: *Backcalculation models describing the key processes determining the pattern of the Bovine Spongiform Encephalopathy (BSE) epidemic in British cattle are derived that allow for infection from feed as well as maternal and direct horizontal transmission. Heterogeneous susceptibility classes are also incorporated* *into the analysis. Maximum likelihood methods are used to estimate parameters and to obtain confidence intervals from available experimental and epidemiological data. A comprehensive sensitivity analysis of model parameters and distributional assumptions is presented. Additional validation is provided by fitting the model to independent data collected in Northern Ireland. Model estimates and predictions based on BSE case data for Great Britain and Northern Ireland, together with their implications, are reviewed, and future research priorities discussed.*

3. Analysis of New Variant CJD Cases in Great Britain: What Can We Say Now? What Will We Be Able to Say Later?

Ferguson, Neil, Affiliation University of Oxford

Address: Wellcome Trust Centre for the Epidemiology of Infectious Disease Department of Zoology University of Oxford South Parks Road Oxford OX1 3PS

Phone: 011 44 1865 281 228

Fax: 011 44 1865 281 241

Email: neil.ferguson@zoology.oxford.ac.uk

Abstract: By September 30, 1997 22 cases of new variant Creutzfeldt-Jakob disease (vCJD) had been confirmed in Great Britain. In March 1996, the Secretary of State for Health announced that 'There remains no scientific proof that BSE can be transmitted to man by beef, but the Committee have concluded that the most likely explanation at present is that these (10 vCJD) cases are linked to exposure to BSE before the introduction of the specified bovine offal ban in 1989.' Since this time, experimental studies have yielded strong scientific evidence linking BSE and vCJD. Backcalculation methods can be used to estimate future case numbers based on these initial cases; however, the lack of knowledge about the incubation period distribution results in high levels of uncertainty. The performance of backcalculation performed at various points throughout the time scale of the epidemic is evaluated under a variety of assumptions consistent with the currently available data.

List of speakers who are nonmembers: Neil Ferguson, Christl Donnelly

Chapter 6

ASA General Methodology (2)

Program Chair: **Sedransk, Nell** Program Chair: **Baggerly, Keith**

6.1 asa.gen.meth.01

Sponsoring Section/Society: General Methodology

Session Slot: 8:30-10:20 Wednesday
Estimated Audience Size: xx-xxx
AudioVisual Request: xxx

Session Title: Mapping: Statistical Challenges for Inference and Visualization

Theme Session: Yes/No
Applied Session: Yes/No

Session Organizer: **Sedransk, Nell** Case Western Reserve University

Address: Department of Statistics Case Western Reserve University 339 Yost Hall Cleveland, Ohio 44106-7054
Phone: 216-368-6941
Fax: 216-368-0252
Email: nxs20@po.cwru.edu or lfs2@po.cwru.edu

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 0 minutes First Speaker - 25 minutes Second Speaker - 25minutes Third Speaker - 25 minutes Fourth Speaker- 25 minutes Floor Discussion - 10 minutes

Session Chair: **TBN**

Address:
Phone:
Fax:
Email:

1. Flexibility in Hierarchical Models of Disease Rates

Waller, Lance A., University of Minnesota

Address: Division of Biostatistics Box 303 Mayo Building 420 Delaware Street SE University of Minnesota Minneapolis, Minnesota 55455-0592
Phone: 612-625-2138
Fax: 612-626-0660
Email: lance@muskie.biostat.umn.edu
Best, Nicky, Imperial College of Medicine at St. Mary's

Abstract: *We consider the impact of different implementations of hierarchical models for disease mapping. We model regional disease counts as Poisson random variables, incorporating covariate information in the log-relative risk. Prior distributions of random effects allow excess heterogeneity and spatial smoothing of rates toward neighboring values. Using simulated data, we assess the impact of prior and hyperprior specifications, and of spatial neighbor definitions. Of particular interest is the accuracy of spatial smoothing in recovering disease patterns due to spatially structured covariates omitted from the model.*

2. Hierarchical Spatial-Temporal Modeling with Geophysical Applications

Berliner, L. Mark, Ohio State University

Address: Department of Statistics Ohio State University 1958 Neil Ave. Columbus, OH 43210-1247

Phone: 614-292-0291
Fax: 614-292-2096
Email: mb@stat.ohio-state.edu
Royle, J. Andy, National Center for Atmospheric Research
Wikle, Christopher K., National Center for Atmospheric Research
Milliff, Ralph, National Center for Atmospheric Research

Abstract: The role of Bayesian hierarchical modeling of spatial-temporal geophysical phenomena is reviewed. An illustrative example, primarily focusing on spatial structure, is presented. In this example a multivariate hierarchical model is developed for wind and pressure fields over the Labrador Sea, based on satellite data. This model is stochastic, but makes use of dynamical relationships within the probability framework. A key feature of this approach is that the relationship between the gradient of pressure and wind components provides the dynamical basis for a stochastic model of the wind, conditional on the pressure field. Furthermore, variations from this gradient relationship are possible due to the stochastic nature of the formulation. Although the methodology can be used without pressure observations, such observations can be included. Extensions to temporal models are discussed.

3. Nonparametric Regression for Geographic Visualization and Analysis of Environmental Policy

Whittaker, Gerald, Economic Research Service, USDA

Address: United States Department of Agriculture Economic Research Service USDA 1800 M Street Room 4177 Washington, DC 20036-5831
Phone: 202-694-5557
Fax: 202-694-5775
Email: gerryw@econ.ag.gov
Scott, David W., Rice University

Abstract: The U.S. Department of Agriculture administers and analyzes a large number of surveys of both economic and environmental data. The data collected have traditionally been *analyzed on the basis of large geographic regions, and results presented in tables. Through the use of the averaged shifted histogram (ASH) approach to nonparametric regression, it is now possible to analyze survey data on the basis of local areas. The results can be presented as surfaces, or maps where the data are geo-referenced. The spatial analysis of covariates using the ASH has provided new insights into policy applications such as the relation of government payments to agricultural land values, spatial distribution of government payments, and use of agricultural nutrients. In a related use of the ASH estimator, we use estimates of surfaces from survey data to link economic and physical models for analysis of environmental issues. Much economic and environmental data are georeferenced to points which are drawn from a spatial distribution. With the ASH estimator data sets which are georeferenced to different sets of points can be linked for analysis. ASH estimates of surfaces representing physical variables allow analysis that links firm-level economic decisions to spatially distributed processes in the physical environment. The economic and physical effects of alternative policies can then be modeled and the results visualized with nonparametric regression. The ASH is very fast compared to most smoothers, and can be modified to account for a complex survey design in estimation. In this paper, the basic ASH methodology is described and several case studies presented.*

4. Recent Developments in Spatial Graphics From Micromaps to Global Grids

Carr, Daniel B., George Mason University

Address: George Mason University Department of Applied and Engineering Statistics George Mason University 9930 Rand Drive Burke, Virgina 22015-3812

Phone: 703-993-1671

Fax: 703-993-1700

Email: dcarr@galaxy.gmu.edu

Olsen, Anthony R., US EPA, Corvallis, OR

Courbois, Jean-Yves (Pip) , Oregon State University, Corvallis, OR

Pierson, Suzanne, OAO Corporation, Corvallis, OR

Kimerling, Jon, Oregon State University

Sahr, Kevin, University of Oregon
White, Denis, US EPA, Corvallis, OR

Abstract: In this talk we focus attention two recent methodological developments for spatial graphics, linked micromap plots (LM plots) and icosahedral Synder equal area (ISEA) global grids. LM plots follow a new template for showing spatially indexed statistics more accurately than traditional choropleth maps. A LM plot consists of parallel sequences of micromap, label and statistical summary panels. The micromaps are often low resolution caricatures that show regions or sites with just enough detail to preserve region recognition and neighbor relationships. The statistical summary panels often take familiar forms such as dot plots, bar plots and box plots, time-series plots and scatterplots. Sorting plus logical or perceptual grouping defines the sequence of panels. Color and position link corresponding graphics elements for each case within a set of panels. The careful design makes LM plots a good device for communicating spatially indexed statistical summaries. Some examples address data sets with millions of observations.

In the last part of the talk, we call attention to ISEA global grids of Jon Kimerly and his colleagues Denis White and Kevin Sahr. The ISEA global grids provide a coherent framework for addressing recurrent issues in environmental sampling and have strong implications for spatial data storage, manipulation, analysis, and presentation. The best known global gridding framework is an equal angular grid that is closely related to the Mercator projection and it's well-known polar distortions. ISEA grids provide global uniformity in terms of cell area. When projected onto an icosahedron, the cells are regular hexagons. While issues remain in regard to changing resolution, ISEA grids may eventually emerge as the preferred framework for integrating data from multiple sources and multiple disciplines.

List of speakers who are nonmembers: None

6.2 asa.gen.meth.02

Sponsoring Section/Society: ASA General Methodology

Session Slot: 8:30-10:20 Tuesday
Estimated Audience Size: xx-xxx
AudioVisual Request: xxx

Session Title: Statistical Investigation Of Gulf War Syndrome

Theme Session: Yes/No
Applied Session: Yes/No

Session Organizer: **Gaver, Donald P.** Naval Postgraduate School

Address: Code OR/GV Naval Postgraduate School Monterey, CA 93943
Phone: 408-656-2605
Fax: 408-656-2595
Email: DGaver@wposmtp.nps.navy.mil

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 5 or 0 minutes First Speaker - 30 minutes (or 25) Second Speaker - 30 minutes Third Speaker - 30 minutes Discussant - 10 minutes (or none) Floor Discusion - 10 minutes (or 5 or 15)

Session Chair: **Gaver, Donald P.** Naval Postgraduate School

Address: Code OR/GV Naval Postgraduate School Monterey, CA 93943
Phone: 408-656-2605
Fax: 408-656-2595
Email: DGaver@wposmtp.nps.navy.mil

1. Fundamental Statistical Questions Arising From Gulf War Syndrome Investigations

Bailar, John C. III, University of Chicago

Address:

Phone:
Fax:
Email:

Abstract: *To establish a physical basis for the Gulf War Syndrome it is necessary to have (1) a precise operational definition of the outcome to be examined, (2) evidence of exposure to the agent in question, and (3) evidence (possibly from non-human studies) that exposure is likely to affect human health. The latter raises issues of the actual levels of exposure, human variation in response (e.g. sensitive persons), the shape of the dose-response relationship, possible interactions with other risk factors, and many other things. To date, (2) has been discussed in extenso, but (1) and (3) have been largely ignored. Without them, no inferences about exposure and health can be considered valid.*

2. The Stress Theory of Gulf War Syndrome: A Political Misstep due to Errors in Measurement

Haley, Robert W., Southwestern Medical School

Address:
Phone:
Fax:
Email:

Abstract: *After treating large numbers of Vietnam veterans with post-traumatic stress disorder (PTSD) resulting from intense combat experiences, mental health professionals expectantly screened large groups of military personnel returning from the 1991 Persian Gulf War (GW) with psychometric PTSD tests, though without the usual psychiatric confirmation. Finding higher, though mostly subdiagnostic, scores in GW veterans than in referents who were not deployed, the authors of 16 published studies concluded that diverse physical symptoms, called GW Syndrome, were due to "mild PTSD," reflecting the small amount and mild nature of combat exposure. This led to a national political consensus that GW Syndrome was a psychological condition due to* *stress and not a physical illness or injury. Correcting the observed PTSD prevalence rates (range 0% to 36%, mean 9%) for measurement errors with previously determined values of the sensitivity (range 0.77 to 0.96) and specificity (range 0.62 to 0.89) of the psychometric tests, however, yields estimated true PTSD rates of 0% for 18 of the 20 reported rates. Mean scores on the Mississippi PTSD Scale in all subgroups of Gulf War veterans were within the range of values for well-adjusted Vietnam veterans (50-89) and far below that of Vietnam veterans with psychiatrically confirmed PTSD (120-140). Most PTSD and "stress-related symptoms" reported in studies of Gulf War veterans appear to represent false-positive errors of measurement reflecting nonspecific symptoms of other conditions. Recently, psychiatrists performing full examinations are finding little to no PTSD, and clinical epidemiologic and animal studies suggest that neurotoxic brain damage from chemical exposures is the explanation.*

3. Health Consequences of Service in the Persian Gulf War: Medical, Policy and Media Syndromes

Louis, Tom, Division of Biostatistics, University of Minnesota

Address: Thomas A. Louis, Division of Biostatistics, University of Minnesota, School of Public Health

Phone:

Fax:

Email:

Abstract: Membership on the Institute of Medicine's Persian Gulf War Health Consequences Committee and attention to media coverage revealed medical, policy and media syndromes. I review evidence on health effects and the steps/missteps of federal agencies and the media. I discuss statistical issues and statistician roles needed for effective participation in such high-profile, high-impact enterprises.

Discussant: **Seglie, Ernest** Office of Sec. of Defense; DOT&E

Address:

Phone:
Fax:
Email:
List of speakers who are nonmembers: None

Chapter 7

ASA Government (2 + 1 competition)

Program Chair: **Doyle, Pat**

7.1 asa.govt.01

Sponsoring Section/Society: Government Statistics Section

Session Slot: 8:30-10:20 Tuesday
Estimated Audience Size: xx-xxx
AudioVisual Request: xxx

Panel Session Title: Greying of the Federal Statistical Workforce–Challenges and Opportunities

The federal statistical workforce is declining in size, as is the rest of the federal government. Concurrently the existing workforce is "aging." The three participants in this session are the heads of the Census Bureau, the National Center for Health Statistics and the Energy Information Administration. Rather than give formal papers, they will have a series of questions provided to them before the session which they will each address. The session topics will include: what are the implications of this "greying" for data products—will we move to the electronic age more rapidly; how do we undertake continuous training with an essentially static workforce; how is the system innovating and adapting; what are the implications of having fewer people and less institutional knowledge to do things the way we have always done them. The session will focus on the ability of the system to adapt—although we are in a static period the system has produced phenomenal change in the way we disseminate information to our customers. How was this accomplished? The participants will also address future plans for adaptation.

Theme Session: Yes
Applied Session: Yes/No

Session Organizer: **Carlson, Lynda** Energy Information Administration

Address: Lynda Carlson, Energy Information Administration, Department of Energy
Phone: 202-426-1068
Fax:
Email: lynda.carlson@eia.doe.gov

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 5 minutes First Speaker - 25minutes Second Speaker - 25 minutes Third Speaker - 25 minutes Fourth speaker - 25 minutes (or none) Floor Discusion - 10 minutes (or 5 or 15)

Session Chair: **Carlson, Lynda** Energy Information Administration

Address: Lynda Carlson, Energy Information Administration, Department of Energy
Phone: 202-426-1068
Fax:
Email: lynda.carlson@eia.doe.gov
Sondik, Edward Director, National Center for Health Statistics

Address:
Phone:
Fax:

Email: efs2@cdc.gov

Hakes, Jay Administrator, Energy Information Administration, Department of Energy

Address:
Phone:
Fax:
Email: jay.hakes@eia.doe.gov

Riche, Martha Director, U.S. Bureau of the Census

Address: U.S. Bureau of the Census, Washington, DC 20233
Phone:
Fax:
Email: mriche@census.gov

Groves, Robert Professor, Joint Program in Survey Methodology, University of Maryland

Address:
Phone:
Fax:
Email: bgroves@survey.umd.edu
List of speakers who are nonmembers: None

7.2 asa.govt.02

Sponsoring Section/Society: Government Statistics Section

Session Slot: 4:00- 5:50 Sunday
Estimated Audience Size: xx-xxx
AudioVisual Request: xxx

Panel Session Title: Changing Our Core Classifications: Challenges and Opportunities

The Federal Statistical System provides data about a number of key variables used by a broad array of users, including Congress, the Administration, state and local governments, and the social science research community. To ensure that data collected by the various decentralized statistical agencies are comparable, the Office of Management and Budget (OMB) develops, regulates and, when necessary, revises standard classifications of such "core" variables as race/ethnicity status, occupation, and industry code to be used in all Federal government surveys. In doing this, OMB, in consultation with users, faces a number of challenging decisions, including: what variables should have standard codes? How often should standard codes be revised? What revisions are needed for specific codes? Since users have differing needs, there are not simple answers to these questions. This panel will describe current practices, challenges, and opportunities exist with regard to developing, regulating, and revising the core classifications and will discuss possible future directions for work in this area.

Theme Session: Yes
Applied Session: Yes/No

Session Organizer: **Shettle, Carolyn** National Science Foundation

Address: National Science Foundation, 4201 Wilson Boulevard, Arlington, VA 22230
Phone: 703-306-1774 x 6906
Fax:
Email: cshettle@nsf.gov

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 5 minutes First Speaker - 30 minutes () Second Speaker - 30 minutes Third Speaker - 30 minutes Discussant - 10 minutes Floor Discusion - 10 minutes

Session Chair: **Wallman, Katherine** Office of Management and Budget

Address: Office of Management and Budget, Washington, D.C. 20603
Phone: 202-395-3093
Fax:
Email: WALLMAN_K@a1.eop.gov
Pinzlerm, Isabel Department of Justice

Address:

Phone:
Fax:
Email:
Gage, Linda California State Finance Department

Address:
Phone:
Fax:
Email:
Pearson, Willie Jr. Wake Forest University

Address:
Phone:
Fax:
Email:
List of speakers who are nonmembers: None

7.3 asa.govt.03

Sponsoring Section/Society: Government Statistics Section ASA

Session Slot: 10:30-12:20 Wednesday
Estimated Audience Size: xx-xxx
AudioVisual Request: xxx

Session Title: The Implications of the North American Industrial Classification System (NAICS) for Official Statistics

For most of this decade, the United States, Canada, and Mexico have worked tointegrate their industrial classification systems, while updating the systems to reflect technological changes and innovations. The North American Industrial Classification System (NAICS) is in place now and government statisticians face a new challenge converting statistical programs to the new system while preserving time series and educating users about NAICS changes. The impact of the changes in converting to NAICS will be enormous, both for the government statisticians making these changes and for the users of the official statistics they produce. The conversion process will take many years to complete and will have unexpected side effects such as redefinition of standard occupation codes. The three speakers in this session were appointed to be their country's official representative for the development of the integrated classifications and for the coordination effort to implement the revised codes.

Theme Session: Yes
Applied Session: Yes

Session Organizer: **Cox, Brenda G.** Mathematica Policy Research Inc.

Address: Brenda G. Cox, Senior Fellow Mathematica Policy Research, Inc. 600 Maryland Ave SW, Suite 550 Washington DC 20024-2512
Phone: 202-484-4231
Fax: 202-863-1763
Email: bcox@mathematica-mpr.com

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 5 minutes First Speaker - 25 minutes Second Speaker - 25 minutes Third Speaker - 25 minutes Discussant - 15 minutes Floor Discussion - 10 minutes

Session Chair: **Cox, Brenda G.** Mathematica Policy Research, Inc.

Address: Brenda G. Cox, Senior Fellow, Mathematica Policy Research, Inc. 600 Maryland Ave SW, Suite 550 Washington DC 20024-2512
Phone: 202-484-4231
Fax: 202-863-1763
Email: bcox@mathematica-mpr.com

1. NAICS and U.S. Statistics

Ambler, Carole, U.S. Bureau of the Census

Address: Chief, Services Division U.S. Bureau of the Census Services Division Room 2633-3 Washington DC 20233
Phone: 301-457-2668
Fax: 301-457-1343
Email: cambler@ccmail.census.gov

Abstract: The North American Industry Classification System (NAICS) was developed in cooperation with Statistics Canada and Mexico's Instituto Nacional de Estadistica, Geografia e Informatica (INEGI) to provide for comparable industrial statistics across the three countries. NAICS is the first ever classification system to be constructed on a consistent conceptual framework. Such a framework will provide consistent statistics across the three countries that can be used for measuring productivity and unit labor costs, constructing input-output tables, and other uses that imply the analysis of production relationships in the economy. This paper discusses the impact that NAICS will have on U.S. industrial statistics and focuses on Bureau of the Census and Bureau of Labor Statistics Programs. Issues such as the redesign of the current services annual program at the Census Bureau and the need to produce time series data are also discussed.

2. NAICS: The Canadian Perspective

Nijhowe, Shaila, Statistics Canada

Address: Director, Standards Division Statistics Canada Classification Systems 11th Floor, Section D3, Jean Talon Building Tunney's Pasture Ottawa, Ontario, CANADA K1A 0T6

Phone: 613-951-8577

Fax: 613-951-8578

Email: nijhsha@statcan.ca

Abstract: The North American Industry Classification System (NAICS) has been developed against the background of the North American Free Trade Agreement (NAFTA) and the need for comparable industry statistics to analyze issues of industrial performance and productivity for which a supply side approach was adopted in its creation. This paper will discuss the impact of the development of NAICS on the annual business statistics program of Statistics Canada in terms of the representation of the structure of the Canadian economy. Issues relating to the conversion of the Business Register, the phased conversion of the annual surveys and the need to produce NAICS based statistical time series will also be discussed.

3. NAICS and Official Statistics in Mexico

Ordaz, Enrique, INEGI, Mexico

Address: Instituto Nacional de Estadistica Geografia e Informatica Torre El Dorado Republica Mexicana esq. Republica de El Salvador Pent House Fracc. El Dorado 20230 Aguascalientes, Ags. MEXICO

Phone: 524-978-7903

Fax: 524-978-7959

Email: eordaz@dge.inegi.gob.mx

Abstract: Jointly developed in cooperation with the statistical agencies of the United States and Canada, the North American Industrial Classification System (NAICS) will provide users with comparable data and a new description of the three countries' economic structure. This paper discusses the issues associated with this integration, as well as the process of NAICS implementation in INEGI's programs of economic census and surveys and in particular treatment of time series.

Discussant: **Werking, George** U.S. Bureau of Labor Statistics

Address: Assistant Commissioner Office of Federal/State Programs U.S. Bureau of Labor Statistics 2 Massachusetts Ave N.E. , Room 4840 Washington, D.C. 20212

Phone: 202-606-6528

Fax: 202-606-6654

Email: Werking_G@BLS.GOV

List of speakers who are nonmembers: : Carole Ambler, Shaila Nijhowe, Enrique Ordaz

Chapter 8

ASA Health Policy Statistics (1 + 2 competition)

Program Chair: **Christiansen, Cindy L.**

8.1 asa.health.pol.01

Health Policy Statistics Section

Session Slot: 8:30-10:20 Monday
Estimated Audience Size: 75-125
AudioVisual Request: None

Session Title: Multiple Measurements/Multiple Levels in Health Care Research

Theme Session: Yes
Applied Session: Yes

Session Organizer: **Christiansen, Cindy L.** Harvard Medical School and Harvard Pilgrim Health Care

Address: Harvard Medical School and Harvard Pilgrim Health Care, 126 Brookline Ave., Suite 200, Boston, MA 02215
Phone: (617) 421-2430

Fax: (617) 859-8112
Email: cindy@hphc.org

Session Timing: 110 minutes total (Sorry about format):

3 speakers and 1 discussant Opening Remarks by Chair - 3 minutes First Speaker - 30 minutes Second Speaker - 30 minutes Third Speaker - 30 minutes Discussant - 10 minutes Floor Discusion - 7 minutes

Session Chair: **McCaffrey, Dan** RAND

Address: RAND, 1700 Main St., PO Box 2138, Santa Monica, CA 90407-2138
Phone: (310) 393-0411 x6735
Fax: (310) 451-7004
Email: daniel_mccaffrey@rand.org

1. Methods for the Analysis of Signal Data in Medicine

Land, Stephanie, Carnegie Mellon University

Address: Department of Statistics Carnegie Mellon University Pittsburgh, PA 15213
Phone: (412) 268-1880
Fax: (412) 268-7828
Email: steph@stat.cmu.edu

Abstract: Medical researchers often need to relate signal data (e.g. analog time series or frequently measured time series) to covariates such as the health or treatment status of patients. This talk will focus on regression estimation and inference for such problems.

2. Analysis of hospital quality monitors using hierarchical time series models

Aguilar, Omar, ISDS Duke University

Address: Duke Univeristy Box 90251, Durham NC, 27708

Phone: 919-684-8840
Fax: 919-684-8594
Email: omar@stat.duke.edu
West, Mike, ISDS Duke University

Abstract: The VA management services department invests considerably in the collection and assessment of data to inform on hospital and care-area specific levels of quality of care. Resulting time series of "quality monitors" provide information relevant to evaluating patterns of variability in hospital-specific quality of care over time and across care areas, and to compare and assess differences across hospitals. In collaboration with the VA management services group we have developed various models for evaluating such patterns of dependencies and combining data across the VA hospital system. This paper provides a brief overview of resulting models, some summary examples on three monitor time series, and discussion of data, modelling and inference issues. This work introduces new models for multivariate non-Gaussian time series. The framework combines cross-sectional, hierarchical models of the population of hospitals with time series structure to allow and measure time-variations in the associated hierarchical model parameters. In the VA study, the within-year components of the model describe patterns of heterogeneity across the population of hospitals and relationships among several such monitors, while the time series components describe patterns of variability through time in hospital-specific effects and their relationships across quality monitors. Additional model components isolate unpredictable aspects of variability in quality monitor outcomes, by hospital and care areas. We discuss model assessment, residual analysis and MCMC algorithms developed to fit these models, which will be of interest in related applications in other socio-economic areas.

3. Analyzing patients' satisfaction with care using multi-level random-effects models

Hui, Siu L., Indiana University School of Medicine

Address: Division of Biostatistics Indiana University School of Medicine Riley Research Wing, 135 699 West Dr. Indianapolis, IN 46202-5119
Phone: 317-274-2661
Fax: 317-274-2686
Email: siuhui@mako.biostat.iupui.edu
Zhou, Xiao-Hua (Andrew), Indiana University School of Medicine
Perkins, Tony, Regenstrief Institute for Health Care
Tierney, Williams M., Regenstrief Institute for Health Care

Abstract: *Since patient satisfaction has become more and more an important indicator of health care quality, in this report, we studied patients' satisfaction among elderly sick patients in an academic outpatient clinic. As part of a study of advance directives in an academic outpatient clinic, we interviewed a cohort of more than a thousand elderly patients, many with significant chronic illness, after scheduled appointments with their primary care physicians. Each patient was followed for one year, and after each scheduled visit, the patient was interviewed by a research assistant who administered two patient satisfaction questionnaires: (1) the ten-item instrument created by the American Board of Internal Medicine (ABIM), which evaluates satisfaction with doctor-patient communication and (2) the seven-item questionnaire from the Medical Outcomes Study (MOS), which assesses satisfaction with the visits just completed. We were interested in identifying clinical factors that might affect patients' satisfaction with care. Due to the hierarchical nature of the data and multidimensional nature of patients' satisfaction with care, we considered covariates at three levels, visit, patient, and doctor. We proposed the use of multi-level random-effects regression models to study effects of three-level covariates on patients' satisfaction scores.*

Discussant: **Katz, Barry** Indiana University School of Medicine

Address: Division of Biostatistics Indiana University School of Medicine Riley Research WING, 135 699 West Dr. Indianapolis, IN 46202-5119
Phone: 317-274-2674

Fax: 317-274-2678
Email: katz@mako.biostat.iupui.edu
List of speakers who are nonmembers: None

8.2 asa.health.pol.02

Sponsoring Section/Society: ASA Health Policy Statistics

Session Slot: 10:30-12:20 Tuesday
Estimated Audience Size: 75-125
AudioVisual Request: None

Session Title: Statistical Tools for Studying Health Care Costs: What Works and Where Do We Need Improvements?

Hillary Clinton, CEO's, drug companies, government officials and patients share concerns over health care costs. This session focuses on several aspects of the problem. Breast cancer screening has been shown to reduce mortality rates in some age groups of women. The first talk considers the cost-effectivenss of targeting certain risk groups for this screening. This talk is followed by one which discusses pitfalls one often encounters when using quality of life years and discounting of later life years in a cost-effective analysis. Examples will be discussed in which the study results (and conclusions) were highly dependent upon the particular weighting method used. The third talk investigates statistical modeling of health care costs. The discussion is applied to the problem of casemix adjustment when setting payments to health plans for their patients. The talks are discussed by Sue Rosenkranz from Harvard Pilgrim Health Care, a large health maintenance organization in the Northeast.

Theme Session: Yes
Applied Session: Yes

Session Organizer: **Christiansen, Cindy L.** Harvard Medical School and Harvard Pilgrim Health Care

Address: Harvard Medical School and Harvard Pilgrim Health Care, 126 Brookline Ave., Suite 200, Boston, MA 02215

Phone: 617-421-2430
Fax: 617-859-8112
Email: cindy@hphc.org

Session Timing: 110 minutes total (Sorry about format):

3 speakers and 1 discussant Opening Remarks by Chair - 3 minutes First Speaker - 30 minutes Second Speaker - 30 minutes Third Speaker - 30 minutes Discussant - 10 minutes Floor Discusion - 7 minutes

Session Chair: **Hilbe, Joe**

Address:
Phone:
Fax:
Email:

1. Breast Cancer Risk and Cost-Effectiveness of Screening Mammography

McGuigan, KA, UCLA School of Public Health, Dept Health Services

Address: RAND, 1700 Main St., PO Box 2138 Santa Monica, CA 90407-2138
Phone: 310-393-0411 x7698
Fax: 310-451-7025
Email: mcguigan@rand.org

Abstract: The cost-effectiveness of a screening program can be facilitated by targeting high-risk populations, as the performance characteristics (positive predictive value) of a screener improve as a function of increased prevalence of disease in a population. Using data from the Breast Cancer Detection Demonstration Project (BCDDP), parametric (logistic regression) and nonparametric (recursive partitioning, artificial neural networks) models are compared to determine which methods better discriminate cancer cases from controls based on known risk factors.

Performance characteristics include model sensitivity, specificity, positive predictive value, and area under the receiver operating characteristic curve (ROC). The best-performing model is used to distinguish low vs normal and high risk groups for alternative screening schedules. Discussion addresses whether current information about epidemiologic risk factors can be used to obtain improved yields with lower overall costs, subject to the constraint of minimizing missed cases (false negatives).

2. *Dangers of Discounting and QALYs*

Palmer, J. Lynn, M.D. Anderson Cancer Center

Address: Department of Biomathematics, 1515 Holcombe Blvd., Houston, TX 77030
Phone: 713-792-7570
Fax: 713-792-4262
Email: jlp@odin.mdacc.tmc.edu

Abstract: Many traditional methods of analysis of health care policy data use weighting methods which may make interpretation of their results difficult. Any analysis using quality of life years as an outcome variable should make explicit in its conclusions how weighting or discounting of early versus late life years was determined, and how its conclusions would have changed if slight adjustments in these weights would have been made. This presentation will include examples of how these difficulties can be (easily) encountered and how results can be clearly presented.

3. *How Well Do Models Work? Predicting Health Care Costs*

Ash, Arlene S., Boston University School of Medicine

Address: Boston University School of Medicine 720 Harrison Ave, #1108, Boston, MA 02118
Phone: 617-638-8188
Fax: 617-638-8026
Email: aash@bu.edu

Abstract: The goal is to pay health plans (such as HMOs) for the expected cost of care for the persons they enroll. Models based on age and sex can reliably distinguish population subgroups whose costs differ by as much as 10 to 1; however, R-squares for age-sex models rarely exceed 0.02. Health-based predictions, such as those produced by Diagnostic Cost Group (DCG) models, explain more of the variation, can detect subgroups whose costs differ by more than 100 to 1 and have larger R-square values, but still, generally less than 0.10. In the context of producing models to be used in setting payments to health plans for their enrollees, we address questions such as: how much better is one model than another? in what way does a model fail? how good is good enough?

Discussant: **Rosenkranz, Susan L.** Harvard Pilgrim Health Care

Address: Deputy Medical Director's Office, Harvard Pilgrim Health Care, BPW-2, 10 Brookline Place West, Brookline, MA 02146
Phone: (617)730-4791
Fax: (617)731-8249
Email: susan_rosenkranz@hphc.org
List of speakers who are nonmembers: None

8.3 asa.health.pol.03

Sponsoring Section/Society: Health Policy Statistics Section

Session Slot: 2:00- 3:50 Wednesday
Estimated Audience Size: 75-125
AudioVisual Request: None

Session Title: Statistical Models and Methods for Health Care Policy Research

Theme Session: Yes
Applied Session: Yes

Session Organizer: **Christiansen, Cindy L.** Harvard Medical School and Harvard Pilgrim Health Care

Address: Harvard Medical School and Harvard Pilgrim Health Care, 126 Brookline Ave., Suite 200, Boston, MA 02215
Phone: (617) 421-2430
Fax: (617) 859-8112
Email: cindy@hphc.org

Session Timing: 110 minutes total (Sorry about format):

3 speakers and 1 discussant Opening Remarks by Chair - 3 minutes First Speaker - 30 minutes Second Speaker - 30 minutes Third Speaker - 30 minutes Discussant - 10 minutes Floor Discusion - 7 minutes

Session Chair: **Bailey, R. Clifton** HCFA

Address:
Phone:
Fax:
Email: clifton.bailey@tcs.wap.org

1. Estimating the Effect on Survival of Detecting Cancers By Screening Using Instrumental Variables: An Example With Breast Cancer Screening

McIntosh, Martin, University of Washington, Department of Biostatistics and Fred Hutchinsen Cancer Research Center

Address: Department of Biostatistics, University of Washington, F-600 Health Sciences 357232, Seattle, Washington 98195
Phone:
Fax: (206)685-1525
Email: mcintosh@biostat.washington.edu

Abstract: Evaluating the effectiveness of cancer screening trials has many technical difficulties. It is common to evaluate a trial by computing a p-value and estimating effect sizes with comparison groups constructed from the subjects treatment assignments, but using only those subjects who developed cancer (the "cancer cohort"); see for example Shapiro (Cancer, 1997),

Shapiro, et. al (JNCI 1982), Chu, et. al (JNCI, 1988) for examples with breast cancer screening). The difficulty of evaluating these data are that (1) the screened group may find cancers earlier than in the unscreened population ("lead time" bias), (2) the screening may find cancers that go undiagnosed in the unscreened population, and (3) some subjects assigned to be screened may have refused.

Each difficulty makes constructing comparison groups difficult: to whom do you compare the screening group cancers? One solution to (1) is to follow subjects until well after the screening has stopped, then use all cancer cases in both groups for comparison. The solution to (2) has been to 'hope' that the long followup time finds the number of cancers in the unscreened group 'caught up' to those in the control group. If catchup does not occurr, then the subjects cancer status depends on the treatment assignment (i.e., it is in the causal pathway), and treating it as a covariate does not lead to valid causal interpretation). Even if catchup does occur, including cancer cases who could not have been detected by screening attenuates the effect size estimate, and leads to substantial bias. Problem (3) has a similar difficulty, because a refuser who was assigned to the control would appear there as a complier, and so compliance staus too is effected by the treatment assignment, and cannot be treated as a covariate.

The methods to deal each of these have been ad-hoc, and admit substantial bias. Here we show how using the Imbens and Rubin (1996a, 1996b) causal framework for dealing with noncompliance can be adapted to analyze cancer screening trials, and gives a very simple and elegant estimate of the causal effect of screening (or screen detection) on patient outcomes (survival time, cancer stage, tumor size, lead time effect). This method deals with all these biases simultaneously.

2. Models for Health Care Planning

Carrier, K.C., University of Alberta

Address: 632 CAB, Department of Mathematical Sciences, Edmonton, Alberta, CANADA, T6G 2G1

Fax: 403-492-6826

Email: kcarrier@fisher.stat.ualberta.ca

Abstract: In health services research, as in any other social science research, the first step in data analysis is to identify the factors deemed relevant for inclusion in the estimation. The next pertinent decisions are the selection of the unit of analysis and an appropriate treatment of the parameters in the equation. In health services research, where health effects are potntially attributable to area level, or ecological variables, there are many critical methodological issues to ponder. All analytic approaches have some risks and some disadvantages. If the analytic results and recommendations were to be implemented successfully in practices of health care planning, it is imperative to understand these risks and disadvantages as well as advantages. Across Canada, health care reform has taken place recently. Using administrative and census data from three Canadian prairie provinces (Manitoba, Alberta, Nova Scotia), we explore weaknesses and strengths of some alternative approaches, with research objectives to efficiently and objectively evaluate the health care reform. We will focus on the relationship between policy changes and changes in overall utilization of hospitals and physicians according to area of residence, area of service (urban and rural) and types of hospital (teaching, urban community, rural) by groups at risk (elderly and those in low socioeconomic conditions).

3. Using Difference (Rather Than Ratio) Measures in Clustered Healthcare Data

Localio, Russell, University of Pennsylvania, Center for Clinical Epidemiology and Biostatistics

Address: Rm 606, Blockley Hall, 423 Guardian Dr., Philadelphia PA, 19104-6021
Fax: 215-573-4865
Email: rlocalio@cceb.upenn.edu
Ten Have, Thomas, University of Pennsylvania, Center for Clinical Epidemiology and Biostatistics
Berlin, Jesse, University of Pennsylvania, Center for Clinical Epidemiology and Biostatistics

Abstract: Much theoretical and applied research on clustered categorical data relies on ratio measures (such as odds ratios) to demonstrate treatment or program effectiveness. These measures, while convenient to the statistician for theoretical reasons, are often less meaningful to the clinical or health policy audience. We investigate the methodological and practical challenges of using measures such as risk difference in the context of clustered health care data. Our example is a two-period cross sectional analysis of the process of cardiovascular care in 197 hospitals (and 23,000 patients) in one state in a program sponsored by the Health Care Financing Administration and its local PRO (Keystone Peer Review, Inc). With multiple endpoints per patient and differing numbers of endpoints per hospital, we also investigate the effects of the number of clusters (hospitals), the number of patients per hospital (cluster size), and the correlation of endpoints within hospital on the choice of estimation techniques. In addition, process endpoints being highly correlated within hospital, we shall address the impact of levels of within-cluster correlation on the choice of estimates.

Discussant: **Gatsonis, Constantine** Brown University

Address:
Phone:
Fax:
Email: gatsonis@jenny.biomed.brown.edu
List of speakers who are nonmembers: None

Chapter 9

ASA Marketing (1)

Program Chair: **Elrod, Terry**

9.1 asa.marketing.01

Sponsoring Section/Society: ASA Section for Statistics in Marketing

Session Slot: 2:00- 3:50 Monday
Estimated Audience Size: 45
AudioVisual Request: none

Session Title: New Methods for Marketing Research

Theme Session: No
Applied Session: No

Session Organizer: **Elrod, Terry** Univ. of Alberta

Address: 3-23 Faculty of Business; Univ. of Alberta; Edmonton, Alberta; Canada T6G 2R6
Phone: 403-492-5884
Fax: 403-492-3325
Email: Terry.Elrod@UAlberta.ca

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 0 minutes First Speaker - 25 minutes Second Speaker - 25 minutes Third Speaker - 25 minutes Fourth Speaker - 25 minutes Floor Discusion - 10 minutes

Session Chair: **Elrod, Terry** Univ. of Alberta

Address: 3-23 Faculty of Business; Univ. of Alberta; Edmonton, Alberta; Canada T6G 2R6
Phone: 403-492-5884
Fax: 403-492-3325
Email: Terry.Elrod@UAlberta.ca

1. Manifest versus Latent Representations in Multiple Category Choice

Russell, Gary J., Univ. of Iowa

Address: Univ. of Iowa; College of Bus. Admin.; 108 Papajohn Bus. Admin. Bldg.; Iowa City, IA 52242-1000
Phone: 319-335-0993
Fax: 319-335-3690
Email: gary-j-russell@uiowa.edu

Abstract: Multiple category choice is a "pick-any" selection task in which consumers purchase a subset of n alternatives. Correlation among individual choices can be induced using autoregressive models defined with observed indicator variables or with latent utilities. We contrast the two approaches using both statistical and marketing theoretic criteria.

2. Modelling Fuzzy Data in Qualitative Marketing Research

Rust, Roland T., Vanderbilt University

Address: Owen Graduate School of Management; Vanderbilt University; 401 21st Avenue South; Nashville, TN 37203
Phone: 615-343-6732
Fax: 615-343-7177
Email: rustrt@ctrvax.vanderbilt.edu

Varki, Sajeev, University of Auckland
Cooil, Bruce, Vanderbilt Univ.

Abstract: In qualitative marketing research the data items to be classified are often fuzzy (belonging to more than one category), which complicates the judgment task. Latent class models, besides being misspecified in the presence of fuzzy data, often fail to converge. We adapt the grade of membership model of Manton, Woodbury, and Tolley to the context of judgment-based classification. This unconditional model is a generalization of the latent class model. We also propose a conditional model that is asymptotically equivalent to the unconditional model, but is much faster.

3. Weak Models Versus Weak Data: Diagnostic Aids for Ambiguous Results

Stewart, David W., University of Southern California

Address: Department of Marketing; Marshall School of Business; University of Southern California; Los Angeles, CA 90089-1421
Phone: 213-740-5037
Fax: 213-740-7828
Email: dstewart@bus.usc.edu

Abstract: Numerous occasions arise on which the results of a model fitting exercise produce ambiguous results. Such ambiguity may arise either because the data are insufficient to provide a strong test of the model or because the model under test is inherently weak. Although the source of ambiguity has important implications, it is not always obvious whether the problem resides in the data or the model. The present paper provides a discussion of this problem and offers two diagnostic procedures for differentiating cases where data are weak from cases where the model itself is weak.

4. Obtaining Product-Market Maps From Pairwise Brand Preferences

Elrod, Terry, Univ. of Alberta

Address: 3-23 Faculty of Business; Univ. of Alberta; Edmonton, Alberta; Canada T6G 2R6
Phone: 403-492-5884
Fax: 403-492-3325
Email: Terry.Elrod@UAlberta.ca

Abstract: This talk introduces product-market maps and explains the advantages of estimating them from consumer preferences for existing brands. A product-market map is fit to pairwise preferences obtained from users of air traffic management systems. An additional analysis of brand perceptions assists in interpretation and verification of the map.

List of speakers who are nonmembers: None

Chapter 10

ASA Physical and Engineering Sciences (5)

Program Chair: **Lin, Dennis K.J.**

10.1 asa.phys.es.01

Sponsoring Section/Society: ASA-SPES

Session Slot: Sunday, 4:00 - 5:50
Estimated Audience Size: 100-150
AudioVisual Request: Two Overheads

Session Title: Data Mining

Theme Session: Yes
Applied Session: Yes

Session Organizer: **Edelstein, Herbert A.** Two Crows Corporation

Address: 10500 Falls Road Potomac, MD 20854
Phone: (301) 983-3555
Fax: (301) 983-2554
Email: herb@twocrows.com

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 5 minutes First Speaker - 35 minutes Second Speaker - 35 minutes Discussant - 15 minutes (or 10 or 20) Floor Discussion - 10 minutes (or 5 or 15)

Session Chair: **Small, Robert D.** Two Crows Corporation and Duke Clinical Research Institute

Address: 2024 West Main Street Durham, NC 27705
Phone: (919) 286-8917
Fax: (919) 286-0570
Email: bob@twocrows.com

1. New Insights from Applying Data Mining to Urology

Tigrani, Vida S., UCSF School of Medicine

Address: UCSF School of Medicine Department of Urology Mountain View, CA
Phone:
Fax:
Email: vtigran@itsa.ucsf.edu
John, George H., Epiphany Marketing Software

Abstract: Data mining is an umbrella term referring to the process of discovering patterns in data, typically with the aid of algorithms to automate part of the search. These methods come from disciplines such as statistics, artificial intelligence, visualization, pattern recognition, and so forth. Researchers analyzing medical data are typically well-versed in basic statistics, primarily experimental design and hypothesis testing, but have not been exposed to data mining. This paper begins by surveying medical data analysis literature and describing the common statistical methods employed by researchers. We then describe some examples of data mining tools that would be useful in analyzing outcomes of medical procedures. Using a case study from urology, we show how standard statistics are applied, and the additional insight that can be gained from the data mining tools. We conclude that there is ample evidence that data mining methods can provide better out

of sample results, or better insights, or perhaps even the same insights and results obtained more quickly, than basic statistical methods, and that therefore data mining methods should become part of the standard medical data analysis toolbox.

2. Dynamic Similarity: Mining Collections of Trajectories

Grossman, Robert, Center for Data Mining, University of Illinois at Chicago

Address:
Phone:
Fax:
Email: Grossman@uic.edu

Abstract: An important challenge in data mining is to extend some of the standard data mining algorithms from data which is static to data which is time varying. Phrased slightly different, the challenge is to extend data mining algorithms from data sets consisting of vectors to data sets consisting of time varying paths or trajectories.

In this talk we survey some of the results in this area related to the similarity problem (find all trajectories which are similar to a fixed trajectory) and the classification problem (classify this trajectory).

We focus in part on time varying data which arises by sampling trajectories from an underlying family of parameterized dynamical systems. This is the dynamic similarity problem and occurs commonly in practice. For example, there are important applications to robotics and aeronautics. We describe a new algorithm to attack this problem and provide experimental data from these domains to demonstrate its scalability on collections ranging in size from 1 GB to 100 GBs.

Discussant: **Edelstein, Herbert A.** Two Crows Corporation

Address: 10500 Falls Road Potomac, MD 20854

Phone: (301) 983-3555
Fax: (301) 983-2554
Email: herb@twocrows.com
List of speakers who are nonmembers: None

10.2 asa.phys.es.02

Sponsoring Section/Society: ASA-SPES

Session Slot: Monday, 8:30 - 10:20
Estimated Audience Size: 100-150
AudioVisual Request: Two Overheads

Panel Session Title: Key Challenges in Today's Industrial Statisticians

Theme Session: Yes
Applied Session: Yes

Session Organizer: **Lin, Dennis K.J.** Penn State University

Address: 314 Beam Building, Penn State University, University Park, PA 16802-1913
Phone: 814-865-0377
Fax: 814-863-2381
Email: DKL5@psu.edu

Session Timing: 110 minutes total (Sorry about format):

> Opening Remarks by Chair - 5 or 0 minutes First Panel Speaker - 15 minutes Second Panel Speaker - 15 minutes Third Panel Speaker - 15 minutes Floor Discusion - 50 minutes

Session Chair: **Vining, G. Geoffrey** University of Florida

Address: University of Florida Department of Statistics Gainesville, FL 32611
Phone: 352-392-1941 x204

Fax:
Email: vining@stat.ufl.edu
Hahn, Gerald GE

Address: GECRD 4C33, P.O. Box 8 Schenectedy, NY 12301
Phone: 518-387-5558
Fax: 518-387-5714
Email: HAHN@CRDGW2.crd.ge.com
Montgomery, Douglas C. Arizona State University

Address: Department of Industrial Engineering Arizona State University Box 875906 Tempe, AZ 85287-5906
Phone: 602-965-3836
Fax: 602-965-8692
Email: atdcm@asuvm.inre.asu.edu
Landwehr, James M. Bell Labs

Address: Bell Laboratories Room 2C-257 700 Mountain Avenue Murray Hill, NJ 07974-0636
Phone: 908-582-7405
Fax: 908-582-3340
Email: jml@bell-labs.com
List of speakers who are nonmembers: None

10.3 asa.phys.es.03

Sponsoring Section/Society: ASA-SPES

Session Slot: 10:30-12:20 Tuesday
Estimated Audience Size: 100-150
AudioVisual Request: Two Overheads

Session Title: Dimension Reduction

Theme Session: Yes
Applied Session: Yes

Session Organizer: **Lin, Dennis K.J.** Penn State University

Address: 314 Beam Building, Penn State University, University Park, PA 16802-1913
Phone: 814-865-0377
Fax: 814-863-2381
Email: DKL5@psu.edu

Session Timing: 110 minutes total (Sorry about format):

First Speaker - 25 minutes Second Speaker - 25 minutes Third Speaker - 25 minutes Fourth Speaker - 25 minutes Floor Discusion - 10 minutes

Session Chair: **Li, Ker-Chau** UCLA

Address: Department of Mathematics, UCLA, LA. CA 90024
Phone: 310-8254897
Fax: 310-2066673
Email: kcli@math.ucla.edu

1. *High Dimensional Data Analysis Via the SIR/PHD Approach*

Li, Ker-Chau, University of California, Los Angeles

Address: Department of Mathematics, UCLA. LA. CA. 90024
Phone: 310-8254897
Fax: 310-2066673
Email: kcli@math.ucla.edu

Abstract: Dimensionality is an issue that can arise in every scientific field. Generally speaking, the difficulty lies on how to visualize a high dimensional function or data set. To investigate this issue, Li(1991, 1992) proposed a dimension-reduction model of the form

$$Y = f(\beta_1' X, \cdots, \beta_k' X, \epsilon)$$

where X denotes the input and Y denotes the output variable. This is general enough to incorporate many regression models as special cases. However, the novel feature about this model is that f is unknown and so is the distribution of the random error ϵ.

The goal is to find the β vectors so that they can be used to reduce the input dimension when k is small. Two simple methods, sliced inverse regression (SIR) and principal Hessian directions (PHD) have been studied.

In this talk, I will give an overview on some recent results about SIR/PHD and describe some ongoing research in various application areas, including censored regression, error-in-regressor, nonlinear confounding, functional data analysis, nonlinear time series, Monte-Carlo and Bayesian computation, and others.

2. A Systematic Approach to the Analysis of Complex Interaction Patterns in 2-Level Factorial Designs

Filliben, James J., National Institute of Standards and Technology

Address: Statistical Engineering Division, NIST North, room 353, Gaithersberg MD 20899
Phone: 301-9752855
Fax: 301-9904127
Email: liu@cam.nist.gov
Li, Ker-Chau, University of California, Los Angeles

Abstract: *Analysis of data from two-level full factorial designs often ends up with a final prediction equation which gives only the significant main effect and interaction terms. When the number of interactions is small, simple and useful interpretation of the equation can then be drawn immediately.*

This article addresses a different situation where the number of significant interactions may be large so that additional efforts are needed in order to sort out the pattern and the relationship between them. In particular, we bring out a class of models where most interactions can be attributed to just one or two (or a very small number of) factors, and conditional on these factors, the models become essentially linear. We offer a strategy for uncovering this structure by Linear Domain Splitting (LDS) whereby a complicated global model is replaced by a series of local domain-specific linear models. We present a recommended methodology

(PHD–principal Hessian direction) for systematically proceeding from the global equation to local split-domain analyses. The net result is that guided tree-structured paths are offered for visiting the source-of-interaction factors in sequence, which appropriately reflects their relative importance and mutual relationship. The final stage modeling is simpler (linear). The quality of the fit can be assessed separately in each region, and the analyst comes away with greater insight as to the sensitivity and robustness of the various factor effects over various regions. Applications in digital electronics testing is illustrated by analyzing a data set collected for studying the conversion error of a digital-to-analog converter.

3. Uncertainty Analysis in Mathematical Modeling via Statistical Dimension Reduction with Application in Groundwater Modeling

Horng, Ming-Jame, Water Resources Bureau, MOEA, ROC

Address: 10F, 41-3, SEC. 3, Hsin-I Rd Taipei, Taiwan, ROC
Phone: 02-7542080
Fax: 02-7542244
Email: mjhorng@wrb.gov.tw
Li, Ker-Chau, University of California, Los Angeles
Yeh, William W-G,

Abstract: *Mathematical modeling is generally applied to describe various physical systems in many scientific fields. A common way of understanding the behavior of a mathematical model is through computer simulation. Such a task often involves model parameters which have to be set properly. Yet some model parameters can be rather hard to determine. One such example is a groundwater system used for predicting water head at a designated well. The crucial model parameters such as the hydraulic conductivity parameters in various aquifer zones can at best be estimated roughly from related geologic and hydrologic information. The uncertainty from such input model parameters often lead to substantial prediction error. This article takes a new approach to uncertainty analysis of mathematical models via a statistical* *dimension reduction method, Sliced Inverse Regression(SIR). It generalizes differential sensitivity analysis and is able to provide a global view on the nonlinear relationship between the output and the input model parameters.*

4. A Three-Way Subclassification Approach to Multiple-Class Discriminant Analysis

Chen, Chun-Houh, Academia Sinica, ROC

Address: Institute of Statistical Science, Academia Sinica Taipei, Taiwan, ROC

Phone:

Fax: (886)-2-7831523

Email: cchen@stat.sinica.edu.tw

Li, Ker-Chau, University of California, Los Angeles

Abstract: Discriminant analysis problems are much harder to handle when the number of classes k is large. In view of this tendency, a strategy based on three-way subclassification is proposed. The original k-class problem is first decomposed into $k(k-1)/2$ *smaller problems, each involving a 3-way classification between a pair of classes A and B, and a combined third class which is just the union of all other classes. Then the results from each subcalssification will be synthesized under a conditional error rate analysis. Our formulation of 3-way subclassification is ideal for exploring the degree of unanimity among different decisions made in subclassification. This advantage is illustrated well in the context of hand-written digit recognition, using the data set of LeCun et al(1989). A large portion of high quality images in the test set can be easily filtered out by 3-way subclassification - they are predicted with an error rate less than* 1%.

List of speakers who are nonmembers: None

10.4 asa.phys.es.04

Sponsoring Section/Society: ASA-SPES

Session Slot: 10:30-12:20 Wednesday
Estimated Audience Size: 60-80
AudioVisual Request: Two Overheads

Session Title: Chemometrics

Theme Session: No
Applied Session: Yes

Session Organizer: **Remund, Kirk** Pacific Northwest National Laboratory

Address: Research Scientist - Statistics Group Pacific Northwest National Laboratory - Battelle
Phone: 509-372-4729
Fax: 509-375-3614
Email: km_remund@pnl.gov

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 5 or 0 minutes First Speaker - 40 minutes Second Speaker - 40 minutes Floor Discussion - 25 minutes

Session Chair: **Remund, Kirk** Pacific Northwest National Laboratory

Address: Senior Research Scientist - Statistics Group Pacific Northwest National Laboratory - Battelle
Phone: 509-372-4729
Fax: 509-375-2604
Email: km_remund@pnl.gov

1. Variable Selection in Chemometrics

Speigelman, Cliff, Texas A&M University

Address: Dept. Statistics Texas A&M University College Station, TX
Phone: 409-845-8887
Fax: 409-845-3144
Email: cliff@stat.tamu.edu

McShane, Mike, Texas A&M University
Cot, Gerry , Texas A&M University

Abstract: Modern chemical instruments frequently generate data with thousands to tens of thousands of variables. In many situations known inputs influence a few of the many variables. Two excellent MATLAB toolboxes have a library of selection techniques for selecting variables from typical chemometric data. The best known is the PLS_Toolbox and another is the Calibration Toolbox. This talk focuses on the methods used in the Calibration Toolbox and theory underpinning those methods. The methods used in this toolbox are easy to understand and work reasonably for several types of chemometric data. We compare our results to those obtained from the highly reputed PLS_Toolbox.

2. Variable Selection for PLS Calibration Models by Genetic Algorithms

Wise, Barry M., Eigenvector Research, Inc.

Address: 830 Wapato Lake Road, Manson, WA, 98831 USA
Phone: 509-687-2022
Fax: 509-687-7033
Email: bmw@eigenvector.com
Gallagher, Neal B., Eigenvector Research, Inc.

Abstract: When developing predictive models, there are often many predictor variables available on which the models may be based. This is particularly true in spectroscopy, where it is common to have thousands of variables from which to choose. Not all variables are equally useful for prediction. Modern regression methods, such as Partial Least Squares or Projection to Latent Structures, reduce the effect of extraneous variables by emphasizing variables which are correlated with each other and the variable to be predicted. In practice, however, extraneous variables seldom end up with regression coefficients that are identically zero. Because of this, these variables contribute to the prediction error on

new samples. Thus, it would be desirable to eliminate these variables from the regression models entirely.

Several approaches to variable selection have been demonstrated. This includes stepwise regression methods, and other methods based on statistical significance tests. Genetic algorithms (GAs) have also been applied to the variable selection problem. In general, the goal of these methods is to determine a set of predictor variables which can be used with Ordinary Least Squares (OLS) to form predictive models. When this method is used, however, the multivariate advantage, i.e. the averaging effect of combining many variables, is lost.

Other variable selection methods have been developed for use with PLS models, such as Global Optimization of Linear Prediction Error (GOLPE) and Interactive Variable Selection (IVS). Both of these methods rely to some extent on cross-validation to determine which variables will be retained. Cross-validation, however, can "misbehave" when an optimization is done with respect to it, particularly when there are many degrees of freedom available to the optimization.

In this work we present a GA based method for selecting predictor variables for PLS models, resulting in models which retain the multivariate advantage. In addition, the method also attempts to avoid the pitfalls of over-fitting data when prediction error is optimized based on cross-validation. This is accomplished by randomization of test sets, and in spectroscopic applications, dividing the spectra into windows several variables wide which must be kept or deleted together. The method is demonstrated using NIR spectra of diesel fuels. The GA method is compared to several other variable selection techniques.

Discussant: **Hardin, James W.** Stata Corporation

Address: Senior Statistician Stata Corporation 702 University Dr. East College Station, TX 77845
Phone: 409-696-4600

Fax: 409-696-4601
Email: tech-support@stata.com
List of speakers who are nonmembers: None

10.5 asa.phys.es.05

Sponsoring Section/Society: ASA-SPES

Session Slot: 8:30-10:20 Thursday
Estimated Audience Size: 60-80
AudioVisual Request: Two Overheads

Session Title: Reliability Models with Incompletely Identified Causes of Failure

Theme Session: No
Applied Session: Yes

Session Organizer: **Yashchin, Emmanuel** IBM

Address: IBM, Thomas J. Watson Research Center Dept. of Math. Sciences, Box 218 Yorktown Heights, NY 10598
Phone: 914-945-1828
Fax: 914-945-3434
Email: YASHCHI@WATSON.IBM.COM

Session Timing: 110 minutes total (Sorry about format):

> 110 total (missing 10 below) First Speaker - 25 minutes (or 25) Second Speaker - 25 minutes Third Speaker - 25 minutes Discussant - 20 minutes Floor Discussion - 5 minutes

Session Chair: **Yashchin, Emmanuel** IBM

Address: IBM, Thomas J. Watson Research Center Dept. of Math. Sciences, Box 218 Yorktown Heights, NY 10598
Phone: 914-945-1828

Fax: 914-945-3434
Email: YASHCHI@WATSON.IBM.COM

1. Survival Analysis With Partially Resolved Failures

Reiser, Benjamin, Univ of Haifa, Israel

Address: Dept. of Statistics, University of Haifa, Haifa 31905,Israel
Phone: 972-4-8240177
Fax: 972-4-8253849
Email: RSST305@UVM.HAIFA.AC.IL
Flehinger, Betty J., IBM Research
Yashchin, Emmanuel, IBM Research

Abstract: *We consider a life testing situation in which systems are subject to failure from independent competing causes. Following a failure, immediate (stage-1) procedures are used in attempt to reach a definitive diagnosis. If these procedures fail to result in a diagnosis, this phenomenon is called masking. Stage-2 procedures, such as failure analysis or autopsy, provide definitive diagnosis for a sample of the masked cases. We show how stage-1 and stage-2 information can be combined to provide statistical inference about (a) survival functions of the individual risks, (b) the proportions of failures associated with individual risks and (c) probability, for a specified masked case, that each of the masked competing risks is responsible for the failure.*

2. Product Design With Incomplete Failure Data

Doganaksoy, Necip, General Electric Corporate R&D

Address: P.O. Box 8, K1-4C43, Schenectady, NY 12301
Phone: 518-387-5319
Fax: 518-387-5714
Email: doganaksoy@crd.ge.com

Abstract: *Proper identification of failure modes (and mechanisms) plays a central role in design of high reliability products. In practical situations, such knowledge and understanding of failure*

modes evolves throughout the product life cycle. It then becomes essential that this information be used properly to guide design changes as well as design of new products. This talk will present some practical issues encountered in utilizing incompletely identified failure data and provide recommendations for research in this area.

3. Cause-Specific Survival Analysis With Covariate-Dependent Failure Type Missingness

Ryan, Louise M., Department of Biostatistics, Harvard School of Public Health, and Dana Farber Cancer Institute

Address: 44 Binney Str., Boston, MA 02115-6084
Phone: 617-632-3602
Fax: 617-632-2444
Email: ryan@jimmy.harvard.edu

Andersen, Janet, Department of Biostatistics, Harvard School of Public Health

Abstract: In studies of cancer and many other chronic disease settings, there is often interest in conducting cause-specific survival analyses which focus on time to death from a particular type of disease, rather than overall survival. Problems arise when failure type information is missing for some individuals in the study. We will discuss methods to adjust for such missingness. We will show that particular care is needed when the missingness probability depends on covariates, and will discuss a model that accommodates this situation. Findings will be illustrated with data from a study in colon cancer where older patients were more likely to have missing failure type information.

Discussant: **Guess, Frank M.** Dept. of Statistics, University of Tennessee

Address: Knoxville, TN 37996-0532
Phone: 423-974-1637
Fax: 423-974-2490
Email: fguess@utk.edu

List of speakers who are nonmembers: None

Chapter 11

ASA Quality and Productivity (3)

Program Chair: **Lagergren, Eric S.**

11.1 asa.qp.01

Sponsoring Section/Society: ASA Quality and Productivity

Session Slot: 2:00- 3:50 Tuesday
Estimated Audience Size: xx-xxx
AudioVisual Request: xxx

Panel Session Title: Lessons Learned From Six Sigma: The Power of Statistical Thinking

Theme Session: No
Applied Session: Yes

Session Organizer: **Hoerl, Roger W.** General Electric Company

Address: General Electric Company Bldg KW-D272 PO Box 8 Schenectady, NY 12301
Phone: (518) 387-4040
Fax: (518) 387-5532

Email: Hoerl@crd.ge.com

This session is intended to communicate the fundamental concepts of the Six Sigma initiative, the results which several corporations have achieved utilizing it, and also to objectively analyze why it has been so successful. The session will begin with a brief introduction to Six Sigma, and then review the experiences of Allied Signal, General Electric, Polaroid, and Motorola. A discussant will then analyze the success of these initiatives from the viewpoint of Statistical Thinking, i.e., identify key attributes of Statistical Thinking which are routinely applied in Six Sigma. Advice will be given to organizations considering implementing Six Sigma. Lastly, time will be allowed for questions from the attendees.

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks - 5 minutes Overview of Six Sigma - 10 minutes Motorola and Polaroid Experience - 20 minutes Allied Experience - 20 minutes GE Experience - 20 minutes Discussant - 20 minutes Floor Discussion - 15 minutes

Session Chair: **Hare, Lynne** Nabisco, Inc.

Address: 200 DeForest Ave. East Hanover, NJ 07936-1944
Phone: 973-503-4154
Fax: 973-503-2968
Email: harel@nabisco.com
Zinkgraf, Steve, Sigma Breakthrough Technologies

Address: Stephen A. Zinkgraf, Ph.D. Sigma Breakthrough Technologies 221 Deer Trail San Marcos, Texas, 78666
Phone: 512-353-7489
Fax: 512-353-7489
Email:
Hill, Bill, Allied Signal Inc

Address: c/o William J. Hill Allied Signal Inc 20 Peabody St Buffalo NY 14210
Phone: 716-827-6273

Fax: 716-827-6373
Email: William.J.Hill@alliedsignal.com
Zinkgraf, Steve, Sigma Breakthrough Technologies
Hoerl, Roger W., General Electric Company

Address: c/o Roger Hoerl General Electric Company Bldg KW-D272 PO Box 8 Schenectady, NY 12301
Phone: 518-387-4040
Fax: 518-387-5532
Email: Hoerl@crd.ge.com
Hahn, Gerald, General Electric Company

Discussant: **Snee, Ron** Bell Atlantic Corporation

Address: Ronald D. Snee Bell Atlantic Corporation 1095 Avenue of the Americas Room 3210 New York, NY 10036
Phone: 212-395-3956
Fax: 212-768-9235
Email: PSMarge@aol.com

11.2 asa.qp.02

Sponsoring Section/Society: ASA Quality and Productivity

Session Slot: 10:30-12:20 Wednesday
Estimated Audience Size: xx-xxx
AudioVisual Request: xxx

Session Title: Planned & Inadvertent Split-Plotting in Industrial Experiments

Theme Session: No
Applied Session: Yes

Session Organizer: **Lucas, James M.** J M Lucas & Associates

Address: 5120 New Kent Rd. Wilmington, DE 19808
Phone: 302-368-1214

Fax: 302-456-4013

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 0 minutes First Speaker - 30 minutes Second Speaker - 30 minutes Third Speaker - 30 minutes Discussant - 10 minutes Floor Discussion - 10 minutes

Session Chair: **Saccucci, Michael S.** Domain Manufacturing Corporation

Address: 3502 Keswick Way Westchester, PA 19382
Phone:
Fax:

1. Split-Plot Response Surface Designs

Lucas, James M., J M Lucas & Associates

Address: 5120 New Kent Rd. Wilmington, DE 19808
Phone: 302-368-1214
Fax: 302-456-4013
Hazel, Malcolm C., Campbell Soup Co.

Abstract: We run a Response Surface Experiment in two different ways; as a (completely) randomized experiment and as a split-plot experiment. Our experiment was motivated by the need to emphasize the difference between a randomized run order and a randomized experiment. The error structure and significance of many of the minor variables changes markedly. The effects of the most significant variables and the overall results of the experiment are similar for both ways of running the Response Surface Experiment because of the dominant effect of a few variables. Analysis procedures are shown which extract all the information from the data. We show how standard regression programs can be used, not only to get the correct tests for Split-Plot Experiments, but also to get good confidence and prediction intervals.

2. Randomizing Run Order Without Resetting Factors

Webb, Derek, Montana St. University

Address: Montana St. University Department of Mathematical Sciences Bozeman, MT 59715

Phone: 406-994-5361

Fax:

Email: webb@math.montana.edu

Lucas, James M., J M Lucas & Associates

Borkowski, John J., Montana St. University

Abstract: Many experiments are run using a random run order, yet if successive runs have the same level of a factor, that factor is not reset. Therefore (complete) randomization is not achieved, and the errors will not be independent. We develop the expected covariance matrix, and calculate the expected bias for significance tests in such experiments. Advantages and disadvantages of using a random run order instead of complete randomization are discussed. In many cases it is appropriate to use a random run order instead of (completely) randomizing the experiment. This is because the maximum variance increases (and the G-efficiency drops) more slowly than the cost drops. This gives justification for the procedure that is so often used in practice. However, it is very important to recognize and understand the difference between a random run order and complete randomization so that the appropriate experimental procedure is used.

Often the experimental goal is to find improved operating conditions or to model how a process works over the experimental region. In many such experimental situations the signal to noise ratio is known to be large. In such situations a random run order is often appropriate. When the purpose of the experiment is scientific understanding (and hypothesis tests), then (complete) randomization is required. When there is a single factor that is hard-to-change then the appropriate split-plot experiment should be considered to get a super-efficient experiment (Anbari and Lucas 1994).

3. Strategies for Designing Complex Split-Plot Experiments

Bancroft, Diccon, W.L. Gore & Associates

Address: 297 Blue Ball Road Elkton, MD 21921
Phone: 302-239-0727
Fax:
Email: dbancrof@wlgore.com

Abstract: Split-plot experiments are blocked experiments in which some experimental factors are applied at the block level. Some strategies for finding efficient split-plot experiments are described. These are particularly useful when the block factors are not orthogonal to subplot treatments. Examples given include a class of strip-plot experiments useful for the estimation of variance components, and an unbalanced factorial split-plot.

Discussant: **Vining, G. Geoffrey** University of Florida

Address: University of Florida Department of Statistics Gainesville, FL 32611
Phone: 352-392-1941 x204
Fax:
Email: vining@stat.ufl.edu

11.3 asa.qp.03

Sponsoring Section/Society: ASA Quality and Productivity

Session Slot: 8:30-10:20 Monday
Estimated Audience Size: xx-xxx
AudioVisual Request: xxx

Session Title: Utilizing Large Industrial Data Sets for Product and Process Quality Improvement

Theme Session: No
Applied Session: Yes

Session Organizer: **Grimshaw, Scott D.** Brigham Young University

Address: Brigham Young University Department of Statistics Provo, UT 84602
Phone: (801) 378-6251
Fax: (801) 378-5722
Email: grimshaw@byu.edu

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 0 minutes First Speaker - 30 minutes Second Speaker - 30 minutes Third Speaker - 30 minutes Discussant - 10 minutes Floor Discusion - 10 minutes

Session Chair: **Grimshaw, Scott D.** Brigham Young University

Address: Brigham Young University Department of Statistics Provo, UT 84602
Phone: (801) 378-6251
Fax: (801) 378-5722
Email: grimshaw@byu.edu

1. Interrogating Large Industrial Databases

MacGregor, John, McMaster University

Address: McMaster University Department of Chemical Engineering 1280 Main Street West Hamilton, Ontario, Canada L8S4L7
Phone: (905) 525-9140 x24951
Fax: (905) 521-1350
Email: macgreg@mcmaster.ca

Abstract: *It is quite common in the process industries for on-line computers to routinely collect measurements every few seconds or minutes on hundreds of process variables. As a result most industries have very large databases containing non-causal data. Fisher's famous statement that "all one can do with such happenstance data is a postmortem to see what they died of" is*

a warning to all those that might try to do too much with such data.

In this presentation we show that with the use of latent variable models and estimation methods based on them (PLS/PCR) one can still extract very useful information from such large multivariate databases for problems such as: (i) multivariate SPC; (ii) building soft (inferential) sensors, (iii) finding process conditions to yield a product with specific properties, or to make the same product in different plants; (iv) establishing multivariate specification limits for product properties; (v) and in certain cases finding directions of steepest ascent for RSM. A key point in the success of these methods is understanding the nature of these databases and how they were generated.

The approach to using latent variable models to interrogate large databases for these problems will be presented and illustrated with some industrial applications.

2. A SAS Procedure for Partial Least Squares and Related Methods

Tobias, Randy, SAS Institute Inc.

Address: SAS Institute Inc. SAS Campus Drive Cary, NC 27513-2414
Phone: (919) 677-8000 x7933
Fax: (919) 677-8123
Email: sasrdt@unx.sas.com

Abstract: Chemometrics, which is the field of statistical analysis for chemical processing, often faces a problem that researchers in other fields would envy—too much data! Typically, instruments monitor and control dozens to hundreds of aspects of a complicated manufacturing process. Analysis must reveal major modes of variation in product quality characteristics, and must identify important factors for predicting and ultimately controlling these characteristics. The method of partial least squares (PLS) is a popular technique for achieving these objectives.

This talk will focus on a new SAS/Stat procedure implementing PLS and related methods. I will discuss the commonalities of

the different procedures I've chosen to implement, give an overview of the syntax and structure of the procedure, and present several examples of its use, in conjunction with other features of SAS software for data manipulation and graphics.

3. Partitions and Hierarchies for SPC with Massive Datasets

Runger, George C., Arizona State University

Address: Arizona State University Industrial and Management Systems Engineering Tempe, AZ 85287-5906
Phone: (602) 965-3193
Fax: (602) 965-8692
Email: runger@asu.edu

Abstract: Increases in information technology continue to increase the size of the datasets and the number of variables available for SPC. An overview of the impact of massive datasets on traditional multivariate SPC will be presented. Specific work on the use of hierarchies and contrasts to decompose the process control problem will be presented. Applications to semiconductor manufacturing and coated products will be made.

Discussant: **Grimshaw, Scott D.** Brigham Young University

Address: Brigham Young University Department of Statistics Provo, UT 84602
Phone: (801) 378-6251
Fax: (801) 378-5722
Email: grimshaw@byu.edu

List of speakers who are nonmembers: None

Chapter 12

ASA Risk Analysis (1)

Program Chair: **Rockette, Howard**

12.1 asa.risk.01

Sponsoring Section/Society: Risk

Session Slot: 10:30-12:20 Wednesday
Estimated Audience Size: 40-60
AudioVisual Request: slide projector, overhead

Session Title: Environmental Risk Assessment

Theme Session: Yes/No
Applied Session: Yes/No

Session Organizer: **Rockette, Howard E.** University of Pittsburgh

Address: Department of Biostatistics 318C Parran Hall Pittsburgh, PA 15261
Phone: 412-624-3022
Fax: 412-624-2183
Email: HERBST+@PITT.EDU

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 0 minutes First Speaker - 30 minutes Second Speaker - 30 minutes Third Speaker - 30 minutes Discussant - 15 minutes Floor Discussion - 5 minutes

Session Chair: **Rockette, Howard E.** University of Pittsburgh

Address: Department of Biostatistics 318C Parran Hall Pittsburgh, PA 15261
Phone: 412-624-3022
Fax: 412-624-2183
Email: HERBST+@PITT.EDU

1. ***Spatial Modeling of Environmental Exposure to Pesticides***
Calvin, James, Texas A&M University

Address: Statistics Department College Station, Texas 77843
Phone: 409-845-3141
Fax: 409-845-3144
Email: CALVIN@PICARD.TAMU.EDU

Abstract: The South Texas Pregnancy Study is a multi-year case-control study of the relationship between pesticides and neural tube defects (NTDs), which are one of the most common forms of birth defects. As part of this study, samples of water, soil and sediment were taken at varying times and at a variety of locations throughout a three county region. Chemical analysis of these samples yields interesting levels of quantities that may be by-products of pesticide use (e.g., atrazine). This talk will discuss the difficulty in modeling these spatially and temporally linked measurements, as well as, techniques for relating the exposure model to NTDs.

2. ***Modeling Reproductive Risks from Environmental Agents***
Mazumdar, Sati, University of Pittsburgh

Address: 306 Parran Hall Department of Biostatistics Pittsburgh, PA 15261

Phone: 412-624-3023
Fax: 412-624-2183
Email: MAZ1@VMS.CIS.PITT.EDU
Mattison, Donald R., University of Pittsburgh
Sussman, Nancy B., University of Pittsburgh

Abstract: Successful reproduction requires complex processes that occur in the male, the female, and the couple. The adverse effects of toxic exposure on these processes can have such impacts on reproductive health as, alternations in sexual behavior, reduced fertility, or adverse pregnancy outcome. Various biomarkers and endpoints can be used to characterize the impact of reproductive toxicity on a population. In this discussion, we present an overview of the statistical methods for the characterization of reproductive toxicity with emphasis on the quantitative estimation of human reproductive risk. We discuss issues related to the temporal effects that may reveal possible recovery of the affected biomarkers, the necessity of utilizing several biomarkers characterizing the same hazard, and the uncertainty in the estimation of the reproductive risk. We describe the usefulness of multivariate mixed-effects models in this endeavor as a useful analytical tool. The need for more andrologic research to gather appropriate data for successful quantitative risk estimation is discussed. Finally, methods and the issues are illustrated by modeling the reproductive risks from 1,2,-dibromo-3, chloropopane and ethylene dibromide to a human population.

3. Increasing Use of Probabilistic Techniques in Human Health Risk Assessments

Sielken, R.L., Sielken, Inc.

Address: 3833 Texas Avenue Suite 230 Bryan, Texas 77802
Phone: 409-846-5175
Fax: 409-846-2671
Email: SielkenINC@aol.com

Abstract: Increasing regulatory demands and new legislation such as the Food Quality Protection Act (FQPA) requiring aggregate exposure assessments combining multiple exposure pathways (e.g., drinking water, multiple foods, and residential) and cumulative risk assessments combining multiple chemicals with a common mechanism of toxicity have stimulated the development and implementation of new probabilistic techniques facilitating higher-tier assessments of exposure incorporating more of the available data than current default initial screening assessments. These new probabilistic techniques are introduced and illustrated in terms of recent submissions of EPA related to pesticides. Real numerical examples illustrate how probabilistic techniques (including Monte Carlo and other distributional analyses) enable the relative likelihood or frequency information in data distributions (e.g., consumption and concentration distributions for drinking water; consumption, application rate, and residue concentration distributions for food sources; joint distributions of several pesticide residues in the same food source; and distributions of residential duration and pesticide application rates for residential exposures) to be used instead of worst case characterizations. Examples include acute, short-term, intermediate-term, and chronic exposures as well as the incorporation of age-dependence and subpopulation differences. Probabilistic techniques make it possible to more realistically combine exposures from multiple time periods, multiple subpopulations, multiple exposure pathways, and multiple chemicals without having to assume to assume the worst case for each component. New probabilistic techniques can help provide greater risk management flexibility and avoid the pitfalls of compounding multiple conservatisms and exaggerating the magnitude of exposure in human health risk assessments.

Discussant: **Wong, Otto** Applied Health Sciences, Inc.; Department of Biostatistics and Epidemiology, Tulane University

Address: 181 Second Avenue Suite 628 Post Office Box 2078 San Mateo, CA 94401

Phone: 650-347-7898

Fax: 650-344-6887

Email: OttoWong@aol.com
List of speakers who are nonmembers: None

Chapter 13

ASA Social Statistics (3)

Program Chair: **Bush, Carolee**

13.1 asa.soc.es.01

Sponsoring Section/Society: ASA-SOC

Session Slot: 10:30-12:20 Thursday
Estimated Audience Size: xx-xxx
AudioVisual Request: xxx

Session Title: Statistical Methodologies for Census 2000

Theme Session: Yes
Applied Session: Yes

Session Organizer: **Thompson, John H.** Bur. of Census

Address: Decennial Census, Washington, DC 20233
Phone: 301/457-3946
Fax: 301/457-3024
Email: john_h_thompson@ccmail.census.gov

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 5 minutes Speaker - 45 minutes
First Discussant - 15 minutes Second Discussant - 15 minutes
Third Discussant - 15 minutes Floor Discussion - 25 minutes

Session Chair: **Thompson, John H.** Bur. of Census

Address: Decennial Census, Washington, DC 20233
Phone: 301/457-3946
Fax: 301/457-3024
Email: john_h_thompson@ccmail.census.gov

1. Statistical Methodologies for Census 2000–Decisions, Issues, and Preliminary Results

Waite, Preston Jay, Bur. of Census

Address: Decennial Census, Washington, DC 20233
Phone: 301/457-3968
Fax: 301/457-3024
Email: pwaite@census.gov
Hogan, Howard, Bur. of Census, DSS

Abstract: Planning and development for Census 2000 is well underway. A fundamental strategy to achieve the goals of increased accuracy at reduced cost is the use of innovative statistical sampling methods as an integral part of census data collection. Census 2000 will be the first decennial census that will use sampling to introduce efficiencies into the follow-up of households that do not mail back a questionnaire. Census 2000 will also include, for the first time, an integrated coverage measurement survey to ensure a complete enumeration by correcting the initial census results for persons who would otherwise be missed. The Census Bureau has been conducting research throughout the decade to develop these statistical methodologies, leading to a complete plan the implementation of these methodologies. The research has been developed and discussed with the a number of advisory and stakeholder groups including the National Academy of Sciences, the American Statistical Society, and the U.S. Congress. As was the

case for previous censuses, the incorporation of sampling as an integral part of census data collection is an extremely controversial topic. A number of issues, both technical and political, have been raised and extensively debated.

This paper will discuss the results of the research that has been conducted this decade and will present the final plan for implementing statistical sampling in Census 2000. The paper will also address the key methodological decisions that have been made in developing the methodologies, and will present some preliminary findings from the 1998 dress rehearsal. Finally, the issues that have been raised will be discussed.

Discussant: **Mulry, Mary** M/A/R/C Research

Address: 7850 North Belt Line Road, Irving, Texas 75063-6098
Phone: 972/506-3971
Fax: 972-506-3504
Email: mary.mulry@marcresearch.com

Discussant: **Rust, Keith F.** Westat, Inc.

Address: 1650 Research Blvd., Rockville, MD 20850-3129
Phone: 301/251-8278
Fax: 301/294-2034
Email: rustk1%westat@mcimail.com

Discussant: **TBN**

Address:
Phone:
Fax:
Email:
List of speakers who are nonmembers: None

13.2 asa.soc.es.02

Sponsoring Section/Society: ASA-SOC

Session Slot: 2:00- 3:50 Wednesday

Estimated Audience Size: xx-xxx
AudioVisual Request: xxx

Panel Session Title: Welfare Reform: How Will We Know What Happened?

On August 22, 1996, President Clinton signed into law the Personal Responsibility and Work Opportunity Reconciliation Act. Three main features of this new law are the elimination of welfare benefits as entitlements, the imposition of work participation requirements, and the introduction of individual lifetime limits on the use of programs funded by TANF (Temporary Assistance for Needy Families).

Participants in this panel discussion are from a variety of backgrounds and thus provide different views of welfare reform: Bill Hobby will chair the session and share his insights on welfare reform in Texas. Dan Weinberg will discuss the Census Bureau's role in ensuring that the capability to measure outcomes associated with welfare reform will be possible, focusing on the new Survey of Program Dynamics. Patricia Ruggles will discuss a research agenda to determine welfare outcomes and her role in working to improve the measurement of child outcomes in state and federal welfare evaluations. Barbara Gault will discuss a research agenda to determine welfare outcomes and the expected impact of welfare on women, particularly single parents. Bill Brister will talk about the Mississippi experience and his role in designing and testing experimental welfare programs in Mississippi. Steve Bell will discuss the initial results of the Urban Institutes AAssessing the New Federalism@ project.

Theme Session: Yes
Applied Session: No

Session Organizer: **Shipp, Stephanie S.** Bur. of Census

Address: Housing and Household Econ. Stats. Div. Washington, DC 20233
Phone: 301/763-8271
Fax: 301/763-8412
Email: stephanie_s_shipp@ccmail.census.gov

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 5 minutes First Panelist - 20 minutes Second Panelist - 20 minutes Third Panelist - 20 minutes Fourth Panelist - 20 minutes Fifth Panelist - 20 minutes Floor Discussion - 15 minutes

Session Chair: **Hobby, William** Rice University

Address: P.O. Box 326, Houston, Texas 77001
Phone: 713/521-3596
Fax: 713/521-3950
Email:
Weinberg, Daniel H. Census Bureau

Address:
Phone:
Fax:
Email:
Ruggles, Patricia Dept. of HHS

Address:
Phone:
Fax:
Email:
Gault, Barbara Institute for Women's Policy

Address:
Phone:
Fax:
Email:
Brister, William Millsaps College

Address:
Phone:
Fax:
Email:
Bell, Stephen Urban Institute

Address:

Phone:
Fax:
Email:
List of speakers who are nonmembers: Hobby, Gault, Brister, Bell

13.3 asa.soc.es.03

Sponsoring Section/Society: ASA-SOC

Session Slot: 8:30-10:20 Monday
Estimated Audience Size: xx-xxx
AudioVisual Request: xxx

Session Title: Public Policy Using Statistcs

Theme Session: Yes
Applied Session: Yes

Session Organizer: **Morton, Sally C.** RAND Corporation

Address: P. O. Box 2138, 1700 Main Street, Santa Monica, CA 90407-2138
Phone: 310/393-0411, X7360
Fax: 310/451-7025
Email: sally_morton@rand.org

Session Organizer: **Rolph, John E.** University of Southern California

Address: Information and Operations Management Dept., Marshall School of Business, Los Angeles, CA 90089-1421
Phone: 213/740-4829
Fax: 213/740-7313
Email: jrolph@rcf.usc.edu

Session Timing: 110 minutes total (Sorry about format):

> Opening Remarks by Chair - 5 minutes First Speaker - 25 minutes Second Speaker - 25 minutes Third Speaker - 25 minutes Discussant - 20 minutes (or none) Floor Discusion - 20 minutes

Session Chair: **Morton, Sally C.** RAND

Address: P. O. Box 2138, 1700 Main Street, Santa Monica, CA 90407-2138, X7360
Phone: 310/393-0411, X7360
Fax: 310/451-7025
Email: sally_morton@rand.org

1. ***Is there Periodicity in the Global Mean Temperature Series?***
Adams, John L., RAND

Address: P. O. Box 2138, 1700 Main Street, Santa Monica, CA 90407-2138
Phone: 310/393-0411, X7917
Fax: 310/451-7025
Email: john_adams@rand.org
Hammitt, James K., Harvard
Hodges, James S., University of Minnesota

Abstract: Periodic fluctuations in global temeratures have drawn interest because of their role in detection and attribution (to natural or anthropogenic sources) of climatic trends. Oscilations with various periods have been reported in the literature. The 2-7 year periodicities have been associated with the EL Nino/Southern Oscillation, but the sources of the "bidecadal" (16-22 year) and longer periodicities are more speculative; possible changes in solar radiance have been suggested. A weakness of earlier research has been the inability to quantify the statistical significance of putative periodicities. We use a non-parametric bootstrap to assess the significance of apparent ENSO-related and bidecadal oscillations in global and hemispheric land and sea temerature series. Our analysis supports the existence of cycles with periods consistent with ENSO but we cannot distinguish the apparent bidecadal oscillations from random variation with data series of the length available.

2. ***Surveying Individuals with Disabilities***
Sampson, Allan R., University of Pittsburgh

Address: Dept. of Statistics, 512 Thackeray Hall Pittsburgh, PA 15260
Phone: 412/624-8372
Fax: 412/361-6042
Email: asampson@stat.pitt.edu

Abstract: *National and local disability programs require accurate and relevant statistics concerning people with disabilities. These data are needed for program advocacy, and implementing and monitoring of new and existing programs. This talk begins with a brief history of disability surveys and a discussion of current frameworks for conceptualizing disability. Issues and problems concerning disability surveys are then examined including reliability, multistage sampling, rare populations, self- identification, and effects of factors that vary over time and environment.*

3. Challenges in Designing and Analyzing Social Experiments

Bell, Robert M., RAND

Address: P.O. Box 2138, 1700 Main Street, Santa Monica, CA 90407-2138
Phone: 310/393-0411, X6672
Fax: 310/451-7025
Email: robert_bell@rand.org
Ellickson, Phyllis L., RAND

Abstract: *Many important public-policy decisions rely on evaluating the effectiveness of social interventions. For example, preventing drug use by adolescents is an important public health and criminal justice priority. One promising approach is school-based prevention programs, which provide a convenient way of reaching almost all adolescents. To choose successfully from among the multitude of available programs, schools need hard evidence about which, if any, prevention strategies work.*

However, such evaluations pose substantial statistical challenges not faced by clinical trials in trying to achieve external and internal validity. This paper describes the design and analysis strategies adopted in a longitudinal experiment conducted at 30 *junior high schools in California and Oregon to evaluate a school-based drug prevention program called Project ALERT. The experiment assessed whether the curriculum reduced initiation and regular use of alcohol, cigarettes, and marijuana by adolescents?*

The paper discusses steps that we took to address the following challenges: achieving full, but realistic implementation of the curriculum; testing the program in diverse environments; obtaining well-matched experimental cells; obtaining reliable measures of use; adjusting for baseline differences among experimental groups; and accounting for intra-school correlation of outcomes.

Discussant: **Mosteller, Frederick** Harvard University

Address: Harvard Univ Science Center 1 Oxford St. Rm 604 Cambridge, MA 02138-2901

Phone: 617-495-2583

Fax: 617-496-8057

Email: NA

List of speakers who are nonmembers: None

Chapter 14

ASA Statistical Computing (5 + 1 competition)

Program Chair: **Wolfinger, Russell D.**

14.1 asa.stat.comp.01

Sponsoring Section/Society: ASA-COMP

Session Slot: 2:00- 3:50 Sunday
Estimated Audience Size: 50-75
AudioVisual Request: None

Session Title: Data Mining

Theme Session: No
Applied Session: Yes

Session Organizer: **Sun, Don X.** Bell Labs, Lucent Technologies

Address: Bell Labs, Lucent Technologies Room 2C-278 Murray Hill, NJ 07974-0636
Phone: 908-582-2149
Fax: 908-582-3340
Email: dxsun@research.bell-labs.com

Session Timing: 110 minutes total (Sorry about format):

First Speaker - 40 minutes Second Speaker - 30 minutes Third Speaker - 30 minutes Floor Discusion - 10 minutes

Session Chair: **Sun, Don X.** Bell Labs, Lucent Technologies

Address: Bell Labs, Lucent Technologies Room 2C-278 Murray Hill, NJ 07974-0636
Phone: 908-582-2149
Fax: 908-582-3340
Email: dxsun@research.bell-labs.com

1. Mining Large Databases in Telecommunications

Buja, Andreas, AT&T Labs - Research

Address: AT&T Labs - Research Room C209 180 Park Avenue P.O. Box 971 Florham Park, NJ 07932-0971, USA
Phone: 973-360-8438
Fax: 973-360-8178
Email: andreas@research.att.com

Abstract: AT&T Labs has built up a datastore of roughly a terabyte of phone call records. In data of this size, the major intellectual effort shifts from modeling to data handling. The reasons are twofold: Traditional statistical systems are not capable of handling data this size, and traditional databases are useless because every analysis step involves a pass over a sizable fraction of the data. To address the demands posed by our datastore, we present a philosphy of small tools that preserves intimacy between users and data. This is in contrast to the philosophy of insulation and abstraction that underlies traditional databases.

Although a large fraction of our efforts is spent handling data and developing tools for handling data, large data also pose interesting statistical problems. For example, the substantive problems (marketing in this case) often do not call for a model of all of the data. Rather, it is sufficient to find and characterize relatively small subsets of interest. This is where datamining is not

just a subarea of statistics: In statistics, we are trained to model all of the data at hand, while in data-"Mining" we try to find worthwhile subsets of the data and leave the rest alone.

2. Data Mining with Extended Symbolic Models

Apte, Chidanand, IBM T. J. Watson Center

Address: Data Abstraction Research IBM Thomas J. Watson Research Center Yorktown Heights, NY 10598
Phone: 914-945-1024
Fax:
Email: apte@watson.ibm.com

Abstract: Symbolic modeling of data with decision trees and decision rules has a certain appeal to data mining application developers. The computationally efficient nature of the modeling methodology, and the inbuilt explanatory nature of the models that are generated, are two often cited reasons for the preferred use of these methods. Traditionally, the applications of these methods had been restricted to classification modeling. Recent extensions to these methods employing ideas from statistics and machine learning have resulted in more general frameworks that continue to exhibit the underlying characteristics but apply to a much wider class of applications. These extended symbolic modeling methodologies permit exciting new application avenues, including regression, probabilistic modeling, non-myopic feature analysis, and integrating data mining into knowledge-based frameworks. A synopsis of work in this area in the data abstraction research group at IBM's T.J. Watson Research Center will be presented.

3. Tips for Data Mining Practitioners

Chu, Robert, SAS Institute Inc.

Address: SAS Institute Inc. Cary, NC 27513
Phone: 919-677-8000
Fax: 919-677-4444
Email: sasrcc@wnt.sas.com

Tideman, Susan, SAS Institute Inc.

Abstract: To get data mining results that improve your enterprise's "bottom line," it sometimes takes more than clean data, good tools and algorithms, and expertise in statistics and/or computing. SAS Institute's experience with the SAS(R) Enterprise Miner(TM) is reflected in a set of tips for getting useful results under various circumstances. This paper, for statisticians and computer professionals, will discuss ten of the tips in detail.

List of speakers who are nonmembers: None

14.2 asa.stat.comp.02

Sponsoring Section/Society: ASA-COMP

Session Slot: 8:30-10:20 Tuesday
Estimated Audience Size: 60-90
AudioVisual Request: None

Session Title: Bootstrapping Time Series

Theme Session: No
Applied Session: Yes

Session Organizer: **Hesterberg, Tim** MathSoft/Statistical Sciences

Address: MathSoft/Statistical Sciences 1700 Westlake Ave. N, Suite 500 Seattle, WA 98109-3044
Phone: (206)283-8802x319
Fax: (206)283-0347
Email: timh@statsci.com

Session Timing: 110 minutes total (Sorry about format):

First Speaker - 35 minutes Second Speaker - 35 minutes Third Speaker - 35 minutes Floor Discusion - 5 minutes

Session Chair: **Hesterberg, Tim** MathSoft/Statistical Sciences

Address: MathSoft/Statistical Sciences 1700 Westlake Ave. N, Suite 500 Seattle, WA 98109-3044
Phone: (206)283-8802x319
Fax: (206)283-0347
Email: timh@statsci.com

1. ***Subsampling with unknown rates of convergence***
Romano, Joseph P., Stanford University

Address: Department of Statistics Stanford University Stanford, CA 94305-4065
Phone: 650-723-6326
Fax: 415-725-8977
Email: romano@stat.Stanford.edu

Abstract: *I will give a brief overview of subsampling, as a general method for the construction of large-sample confidence regions for a general unknown parameter θ associated with the probability distribution generating a stationary sequence $X_1, \ldots, X_n$. In particular, subsampling can work when the bootstrap fails and the general theory supports this statement. The subsampling methodology hinges on approximating the large-sample distribution of a statistic $T_n = T_n(X_1, \ldots, X_n)$ that is consistent for θ at some known rate τ_n.*

Although subsampling has been shown to yield confidence regions for θ of asymptotically correct coverage under very weak assumptions, the application of the methodology as it has been presented so far is liminted if the rate of convergence τ_n happens to be unknown or intractable in a particular setting. In this talk, I will discuss how it is possible to circumvent this limitation by (a) using the subsampling methodology to derivate a consistent estimator of the rate τ_n, and (b) employing the estimated rate to construct asymptotically correct confidence regions for θ based on subsampling.

2. ***Matched–Block Bootstrap for Dependent Data***

Kuensch, Hans-Ruedi, Seminar fur Statistik, ETH Zurich

Address: Seminar fur Statistik ETH Zentrum CH-8092 Zurich Switzerland
Phone: +41-1-632 3416
Fax: +41-1-632 1086
Email: kuensch@stat.math.ethz.ch

Abstract: *The block bootstrap for time series consists in randomly resampling blocks of consecutive values of the given data and aligning these blocks into a bootstrap sample. Hence in the bootstrap sample, different blocks are independent whereas in the original sample they are not. This impairs the performance of the blockwise bootstrap, particularaly in cases where the dependence is strong. Here we suggest a solution ot this problem by aligning with higher likelihood those blocks which match at their ends. This is achieved by resampling the blocks according to a Markov chain whose transitions depend on the data. The matching algorithms we propose take some of the dependence structure of the data into account. They are based on a kernel estimate of the conditional lag one distribution or on a fitted autoregression of small order. Numerical and theoretical analyses in the case of estimating the variance of the sample mean show that matching reduces bias and, perhaps unexpectedly, has relatively little effect on variance. Our theory extends to the case of smooth functions of a vector mean.*

3. Estimation and Prediction for Time Series Using the Bootstrap

Thombs, Lori, University of South Carolina

Address: Department of Statistics University of South Carolina Columbia, SC 29208-0001
Phone:
Fax:
Email: thombs@math.sc.edu

Abstract: *We discuss two problems in time series analysis in which the bootstrap proves to be a useful tool. It is a well known result that if a process is pure white noise, then the autocorrelation estimates, scaled by the square root of the sample size, are asymptotically standard normal. We show that if the underlying process is assumed to be uncorrelated rather than independent, the asymptotic distribution is not necessarily standard normal. We present general asymptotic theory for the estimated autocorrelation function and discuss testing for lack of correlation without the further assumption of independence. Appropriate resampling methods are proposed which can be used to approximate the sampling distribution of the autocorrelation estimates under weak assumptions.*

In a second application, the nonparametric bootstrap is applied to the problem of prediction in autoregresion. Standard prediction techniques for Gaussian time series utilize the result that the conditional distribution of Y[t+k], given the data, is Gaussian as well. As a nonparametric alternative, the bootstrap can be used to estimate the conditional distribution of Y[t+k]. The bootstrap replicated, being generated backward in time, all hav ethe same conditionally fixed p-values at the end of the series. Simulations are presented which illustrate the performance of the proposed bootstrap prediction technique.

List of speakers who are nonmembers: None

14.3 asa.stat.comp.03

Sponsoring Section/Society: ASA-COMP

Session Slot: 2:00- 3:50 Wednesday
Estimated Audience Size: 100
AudioVisual Request: None

Session Title: Internet Developments

Theme Session: No

Applied Session: Yes

Session Organizer: **Narasimhan, Balasubramanian** Stanford University

Address: Department of Statistics Stanford University Stanford, CA 94305
Phone: 650-725-6163
Fax: 650-725-8977
Email: naras@stat.stanford.edu

Session Timing: 110 minutes total (Sorry about format):

First Speaker - 25 minutes Second Speaker - 25 minutes Third Speaker - 25 minutes Fourth Speaker - 25 minutes Discussant - 10 minutes

Session Chair: **Narasimhan, Balasubramanian** Stanford University

Address: Department of Statistics Stanford University Stanford, CA 94305
Phone: 650-725-6163
Fax: 650-725-8977
Email: naras@stat.stanford.edu

1. SticiGui: Statistics Toolbox for Internet and Classroom Instruction with a Graphical User Interface

Stark, P.B., University of California, Berkeley

Address: Deptartment of Statistics University of California Berkeley, CA 94720-3860
Phone: 510-642-1430
Fax: 510-642-7892
Email: stark@stat.Berkeley.EDU

Abstract: SticiGui is a collection of online course material, including a hypertext with links to a glossary, embedded Java applets to illustrate statistical concepts and analyze data, online

interactive lab assignments, automated assignment grading, and logging and querying of student scores to a web site. The material is introductory and largely non-mathematical, but includes case studies in which statistics has been involved in litigation or legislation, and a number of real data sets. For a preview, visit www.berkeley.edu/ stark/Teach/SticiGui97/index.htm . The content is still rough and by no means finished. I will demonstrate the material, and relate some student reactions and some of my experiences using this online approach to teach large (250-350 student) lower-division classes.

2. ***WebStat: What's New in Version 2!***

Ogden, R. Todd, University of South Carolina

Address: Department of Statistics University of South Carolina Columbia, SC 29208
Phone: 803-777-7800
Fax: 803-777-4048
Email: ogden@stat.sc.edu
West, R. Webster, University of South Carolina

Abstract: Introduced in the spring of 1997, WebStat is a complete statistical package written in the form of a Java applet for access over the World Wide Web (WWW) or as a stand-alone application. The package has been used as the primary analysis package in several courses around the world. A brief history of the development of WebStat is given, along with the developers' responses to user feedback. The current and future capabilities of WebStat will be dicussed.

3. ***Introductory Statistics on the Internet via PHP***

de Leeuw, Jan, University of California at Los Angeles

Address: Department of Statistics University of California at Los Angeles 8142 Math Sciences Bldg, Box 951554, Los Angeles, CA 90095-1554
Phone: 310-825-9550
Fax: 310-206-5658

Email: deleeuw@stat.ucla.edu

Abstract: On http://www.stat.ucla.edu/calculators/cdf, there is a calculator that replaces classical statistical tables. In fact, it can do much more. We have calculator pages for the cumulative distribution, for the quantile function, for the probability density/probability mass function. These pages are available for normal, logistic, cauchy, laplace, student, noncentral student, exponential, gamma, chi square, noncentral chi square, f, noncentral f, weibull, beta, uniform, binomial, poisson, negative binomial, discrete uniform, and hypergeometric distributions.

In addition, the calculator can make either point, line, or spike plots of both the cdf and the pdf. Finally, it can sample an arbitrary number of random numbers from the distributions, and email them to the user.

Technical appendix: the calculator uses the C code from Barry Brown et al's dcdflib and randlib. On top of this C code, functions were written to integrate the statistics into PHP/FI, a WWW scripting language. PHP/FI, thus extended, was linked as an Apache module into the WWW server. Thus the statistical functions can be used directly on the html pages. If you want to know more about implementation, see http://www.stat.ucla.edu/-journals/jss/v02/i01.

4. Distributed Computing for Data Analysis

Lang, Duncan Temple, Bell Labs, Lucent Technologies

Address: Bell Labs, Lucent Technologies 700 Mountain Avenue, Room 2C-259 Murray Hill, NJ 07974-2070
Phone: 908-582-3217
Fax: 908-582-3340
Email: duncan@research.bell-labs.com
Hansen, Mark, Bell Labs
Chambers, John, Bell Labs

Abstract: *Statistical computing is part of a more general process, which can be called* computing with data*. Besides traditional statistical analysis, this involves acquiring, organizing, and visualizing data, often in large, structured datasets organized in database management systems and used for purposes beyond analysis. An important challenge for statistical computing (and statistics in general) is to increase the scope of our involvement in this diverse environment. At the same time, the computing environment itself is becoming more diverse in all respects: data and users are widely spread and using many different systems.*

We describe research looking towards the next generation of software for such applications, centered on the idea of distributed computing with data*. By this we mean distributed in two fundamentally different, but related, senses. First, the data and the tasks users apply to the data are distributed geographically, over a heterogeneous network of computers and operating systems. Second, the programming environment we envision is distributed over a variety of languages and other software.*

The system we are designing is centered on the idea of communication of tasks, data description, and application implementation to support distribution in both senses. The support for communication is provided by existing software (CORBA in one trial project, e.g.). The data-analytic programming and, especially, the structure of the relevant data objects, make use of the latest version of S. By combining the features of such software, we expect to provide user/programmers with a more flexible and extensible environment, with many more choices for implementing new ideas. This allows programmers to develop specialized applications in their chosen language and have these provide services to and receive services from other applications locally or via the web. In addition, we anticipate important opportunities to automate much of the data and method definition, exploiting the self-describing nature of classes and methods in S and IDL.

Discussant: **Rosenberger, James L.** Penn State University

Address: Department of Statistics 326 Thomas Building Penn State University University Park, PA 16802-2111

Phone: 814-865-1348
Fax: 814-863-7114
Email: jlr@stat.psu.edu
List of speakers who are nonmembers: None

14.4 asa.stat.comp.04

Sponsoring Section/Society: ASA-COMP

Session Slot: 10:30-12:20 Monday
Estimated Audience Size: 50-75
AudioVisual Request: None

Session Title: Computing for Large Mixed Models

Theme Session: No
Applied Session: Yes

Session Organizer: **Wolfinger, Russell** SAS Institute Inc.

Address: SAS Institute Inc. R-52 SAS Campus Drive Cary, NC 27513
Phone: 919-677-8000 x6695
Fax: 919-677-4444
Email: sasrdw@unx.sas.com

Session Timing: 110 minutes total (Sorry about format):

First Speaker - 35 minutes Second Speaker - 35 minutes Third Speaker - 35 minutes Floor Discusion - 5 minutes

Session Chair: **Wolfinger, Russell** SAS Institute Inc.

Address: SAS Institute Inc. R-52 SAS Campus Drive Cary, NC 27513
Phone: 919-677-8000 x6695
Fax: 919-677-4444
Email: sasrdw@unx.sas.com

1. Making REML Feasible For Mixed-Model Inference From Large Data Sets: Use of the Gibbs Sampler

Harville, David, IBM Thomas J. Watson Research Center

Address: Mathematical Sciences Department, IBM Thomas J. Watson Research Center P. O. Box 218, Yorktown Heights, NY 10598-0218
Phone:
Fax:
Email: harville@watson.ibm.com

Abstract: The computations encountered in the REML (restricted maximum likelihood) estimation of variance components involve the repeated (for each of a number of trial values of the variance ratios) evaluation of the trace and of various other linear combinations of the inverse, say V, of a positive definite matrix W, whose dimensions equal the number of random effects. The matrix W is the matrix obtained from the so-called mixed-model equations by absorbing the equations for the fixed effects into those for the random effects. Typically, W is sparse or exhibits other potentially useful structure. Such structure can be exploited by reexpressing the quantities to be evaluated as expected values of a random vector, say x, whose distribution is multivariate normal with variance-covariance matrix V (and with a null mean vector). The evaluation can then be carried out by an MCMC approach in which the Gibbs sampler is used to draw from the distribution of x. This approach can produce large savings in (computer) time and memory, thereby making REML estimation feasible for much larger numbers of random effects (and consequently for much larger data sets) than ever before.

2. Fast Computing of REML Estimates in Large Linear Mixed Models with Simple Random Effects

Bulavsky, Yurii, Institute of Computational Mathematics and Mathematical Geophysics, Novosibirsk, Russia

Address: Institute of Computational Mathematics and Mathematical Geophysics Novosibirsk Lavrent'ev Ave., 6 630090, Russia

Phone: 007-3832-35-67-21
Fax: 007-3832-32-42-59
Email: bul@osmf.sscc.ru

Abstract: *Considering some effects in a linear model to be random is often a useful tool for modeling uncertainty; however, additional computational cost can be incurred in estimating variance components. A direct implementation of the REML approach involves solving a linear system in the mixed model matrix and computing the inverse of this matrix in each iteration. If the number of levels of random effects is large then forming and directly inverting this matrix may take too much memory and time.*

We consider techniques that avoid the formation and direct inversion of the mixed model matrix, thus making computations much faster. The suggested approach is based on linear algebra formulae for the inverse of an augmented matrix and for the inverse of a low rank modification. We also use some averaging approximations based on the missing-at-random assumption. This leads to approximate but very close to exact REML estimates, and for the test problems considered the differences between the exact and approximate estimates were negligible compared to confidence intervals. The new approach can be tens and hundreds times faster than traditional computations based on directly inverting the mixed model matrix.

3. Direct and Sparse-Matrix Methods of Computing for Mixed Linear Models

Smith, Stephen P., Rodgers Land and Development Company

Address: Rodgers Land and Development Company 5390 Washington St. Napa, CA 94558
Phone: 510-680-6882
Fax:
Email: hucklebird@aol.com

Abstract: *There are three popular approaches to directly compute estimates and predictions of effects in mixed linear models.*

The first approach, the QR algorithm, is suitable for rectangular design matrices. It is promoted mainly for least squares analysis, but the QR algorithm is easily adapted for general mixed models. The second and most common approach is to apply Gaussian elimination or the Cholesky decomposition to the normal equations or the mixed model equations. The third approach is less common, and involves special matrix operations that are suitable for symmetric and indefinite systems of linear equations, called the augmented equations. Because the augmented coefficient matrix is indefinite, it may be necessary to use the Bunch and Parlett factorization to secure numerical stability. The Cholesky decomposition exists for selected row and column permutations of the augmented coefficient matrix, but Cholesky's algorithm may show numerical instability due to the indefiniteness of the coefficient matrix. All of these direct approaches have been adapted to treat sparse matrices, and this enhances their applicability for large mixed models, where computing resources (space and time) are issues. Interestingly, likelihood evaluation is a trivial by-product of prediction and estimation. Hence, variance component estimation is possible via restricted maximum likelihood, by maximizing the objective function that is assembled from the intermediate calculations that result from one of the three approaches. First and second derivatives are useful to optimize the objective function, but these are available via automatic differentiation (forward and backward derivatives). The differentiation of the Cholesky algorithm is a useful result not only for the second approach, but also for the first because the QR algorithm provides an alternative method to compute the Cholesky decomposition. The differentiation of the Bunch and Parlett factorization is also useful, but this needs development.

List of speakers who are nonmembers: None

14.5 asa.stat.comp.05

Sponsoring Section/Society: ASA-COMP

Session Slot: 8:30-10:20 Thursday
Estimated Audience Size: 40-60
AudioVisual Request: None

Session Title: Computer Algebra Systems in a Course in Probability and Statistics

Theme Session: No
Applied Session: Yes

Session Organizer: **Kinney, John** Rose-Hulman Institute of Technology

Address: Rose-Hulman Institute of Technology 5500 Wabash Avenue Terre Haute, IN 47803
Phone: 812-877-8390
Fax: 812-877-3198
Email: john.kinney@rose-hulman.edu

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 5 or 0 minutes First Speaker - 30 minutes (or 25) Second Speaker - 30 minutes Third Speaker - 30 minutes Discussant - 10 minutes (or none) Floor Discusion - 10 minutes (or 5 or 15)

Session Chair: **Kinney, John** Rose-Hulman Institute of Technology

Address: Rose-Hulman Institute of Technology 5500 Wabash Avenue Terre Haute, IN 47803
Phone: 812-877-8390
Fax: 812-877-3198
Email: john.kinney@rose-hulman.edu

1. MAPLE Integrated Into the Instruction of Probability and Statistics

Tanis, Eliot, Hope College

Address: Department of Mathematics Hope College Holland, MI 49422-9000

Phone: 616-395-7527
Fax: 616-395-7123
Email: tanis@math.hope.edu

Abstract: You can use a Computer Algebra System (CAS) such as MAPLE to enhance instruction in probability and statistics. Change parameters of various probability models to illustrate graphically their significance. Make figures that are static in a book come alive using animation. Use the integration or summation capabilities of a CAS to solve interesting problems. Simulate random phenomena to illustrate theoretical concepts. Let a CAS find and graph the p.d.f.'s of order statistics. Use a CAS to simulate the first m (only) out of n order statistics. Involve the students.

2. Mathematica as an Aid in Teaching Probability and Statistics

Kinney, John, Rose-Hulman Institute of Technology

Address: Rose-Hulman Institute of Technology 5500 Wabash Avenue Terre Haute, IN 47803
Phone: 812-877-8390
Fax: 812-877-3198
Email: john.kinney@rose-hulman.edu

Abstract: Examples are given in the use of Mathematica in introductory courses in probability and statistics. Mathematica, as well as other computer algebra systems, allows the instructor to make probability and statistics more visual than has been possible in the past. Calculations are also eased, allowing a de-emphasis on approximations to probability distributions. Classroom tested examples are given. Examples include; probability distributions and their graphs; central limit theorem; bivariate normal distributions; generating functions, estimation problems and distributions of sample statistics.

3. GAUSS; a Fast, Flexible, Comprehensive Programming Language for Scientists, Engineers, and Statisticians

Heller, Barbara, Illinois Institute of Technology

Address: Department of Pharm. and Phys. Sciences University of Chicago, Abbott Hall 947 E. 58th Street Chicago, IL 60637

Phone: 773-834-0378

Fax:

Email: bheller@midway.uchicago.edu

Abstract: The best way to describe GAUSS is that it is a fast and flexible interactive comprehensive programming language. It is the solution for those who do not have the time or the inclination to program in C, FORTRAN, or other languages since it is easy to learn and at the same time flexible and powerful for sophisticated analyses. On the one hand GAUSS is exceptionally fast, providing performance comparable to compiled C or FORTRAN programs. On the other hand it saves you the trouble of progamming from scratch because it contains a large number of built-in procedures such as mathematical and statistical function evaluation, differentiation and integration, linear algbra operations, eigenvalues, Fourier transforms, and random number generation. Also included are GAUSS Applications programs for various specific uses. Of particular interest is a module for maximum likelihood estimation along with an additional set of procedures for statistical inference. Another module of interest is CurveFit which includes procedures for the nonlinear least squares fitting of data with or without weighting. Additional procedures provide for computing bootstrap estimates and distributions. GAUSS also generates 2-D and 3-D publication quality graphs.

List of speakers who are nonmembers: None

14.6 asa.stat.comp.06

ASA Competition: Statistical Computing Section

Session Slot: 10:30-12:20 Wednesday
Estimated Audience Size: 100
AudioVisual Request: None

Session Title: Smoothing Methods and Data Analysis

Theme Session: No
Applied Session: Yes

Session Organizer: **O'Connell, Michael** Becton Dickinson

Address: Becton Dickinson Research Center Research Triangle Park, NC 27709-201
Phone: 919-990-2136
Fax:
Email: Michael_A_OConnell@bdits.bd.com

Session Timing: 110 minutes total (Sorry about format):

TBA (110 minutes total) Opening Remarks by Chair - 5 or 0 minutes First Speaker - 30 minutes (or 25) Second Speaker - 30 minutes Third Speaker - 30 minutes Discussant - 10 minutes (or none) Floor Discusion - 10 minutes (or 5 or 15)

Session Chair: **O'Connell, Michael** Becton Dickinson

Address: Becton Dickinson Research Center Research Triangle Park, NC 27709-201
Phone: 919-990-2136
Fax:
Email: Michael_A_OConnell@bdits.bd.com

1. When is a Feature Really There? The SiZer Approach

Marron, J.S., University of North Carolina

Address: Department of Statistics University of North Carolina Chapel Hill, NC 27599-3260
Phone: 919-962-2188
Fax: 919-962-1279

Email: marron@stat.unc.edu
Chaudhuri, Probal, Indian Statistical Institute

Abstract: When applying smoothing methods to data, the central question is often "which observed features are part of the true underlying curve, and which are spurious sampling artifacts?" An answer is provided using a scale space viewpoint, where one considers the entire family of smooths indexed by the bandwidth. Different bandwidths represent different "levels of resolution of the data". Significance of features that are characterized by Zero crossings of the derivative (such as bumps) is easily understood, assessed and presented, simultaneously over all resolutions, via a color map in scale space.

2. FUNFITS: Fitting Functions to Data

Nychka, Douglas, National Center for Atmospheric Research and North Carolina State University

Address: Geophysical Statistics Project National Center for Atmospheric Research Climate and Global Dynamics Division Boulder, CO 80307
Phone: 303-497-1711
Fax: 303-497-1333
Email: nychka@ucar.edu

Abstract: FUNFITS is a suite of functions that enhance the S-PLUS statistical package and facilitate curve and surface fitting. This software grew up as a project to augment S-PLUS with additional nonparametric fitting strategies. Part of this effort was to focus on three areas rich in applications of function estimation and inference: spatial statistics, response surface methodology and nonlinear time series/dynamical systems. This activity has also lead to methods for space filling designs, appropriate when the response is expected to be a complex surface. One unique feature of FUNFITS is the integration of thin plate spline methodology with the more general estimators based on spatial processes. This functionality is implemented completely in the S language and so

the algorithms and source code are easily accessible. The core methods implemented in FUNFITS are object driven, including prediction, plot and summary functions that help the user interpret results and visualize the estimate.

3. Some Applications of Spline Smoothing

Gu, Chong, National Institute of Statistical Sciences

Address: National Institute of Statistical Sciences PO Box 14006 Research Triangle Park, NC 27709-4006
Phone: 919-685-9323
Fax: 919-685-9310
Email: gu@hp6.niss.org

Abstract: *Nonparametric function estimation procedures are important tools in modern data analysis, of which spline smoothing is among the more effective ones available, especially for multivariate problems. In this talk, we will present some recent developments of spline smoothing in regression, density estimation, and conditional density estimation, and illustrate their applications in data analysis.*

List of speakers who are nonmembers: None

Chapter 15

ASA Statistical Consulting (3)

Program Chair: **LaVange, Lisa M.**

15.1 asa.stat.consult.01

ASA-Statistical Consulting

Session Slot: 10:30-12:20 Monday
Estimated Audience Size: 200
AudioVisual Request: One Overhead Projector

Session Title: Practical Power Analyses for Statistical Consulting

Theme Session: No
Applied Session: Yes

Session Organizer: **Glueck, Deborah H.**

Address: Room N-101A University of Medicine and Dentistry of New Jersey Robert Wood Johnson Medical School 675 Hoes Lane Piscataway, New Jersey 08854-5635
Phone: 908-235-5267
Fax: 908-235-4569
Email: glueckdh@umdnj.edu

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 5 minutes First Speaker - 25 minutes Second Speaker - 25 minutes Third Speaker - 25 minutes Fourth - 25 minutes Floor Discusion - 5 minutes

Session Chair: **Muller, Keith E.** University of North Carolina-Chapel Hill

Address: Department of Biostatistics McGavran-Greenberg, CB#7400 University of North Carolina Chapel Hill, NC 27599
Phone: 919-966-7272
Fax: 919-966-3804
Email: keith_muller@unc.edu

1. Power and Sample Size for Clinical Trials of Survival
Shuster, Jonathan J., University of Florida

Address: Director, Pediatric Oncology Group Statistical Office University of Florida 104 N. Main St. #600 Gainesville, FL 32601
Phone: 352-392-5198
Fax: 352-392-8162
Email: jon@pog.ufl.edu

Abstract: This paper will be devoted to methodology for computing sample size requirements for randomized two-arm clinical trials, whose endpoints are times until an event, such as survival or event-free survival. The methodology presumes that patients enter according to a Poisson Process, and that proportional hazards can be assumed.

Three scenarios will be considered. First, we shall derive the sample size for the situation where no sequential monitoring is utilized. This is covered in the author's book, "Practical Handbook of Sample Size Guidelines for Clinical Trials" (CRC Press 1992). Two sequential designs will utilize an O'Brien-Fleming approach. We shall derive the required properties where continuous sequential monitoring is feasible, thanks to advances in information transfer. Finally, we shall consider a group sequential approach, which is the first to directly plan for outcome-based *timing of looks on the information scale, from the onset of the trial. (Lan and DeMets have looked at the robustness of changing the frequency of looks during the trial, when you have a Z-value within 20% of the O'Brien-Fleming bound. See Biometrics 45, pp. 1017-1020, 1989). At each look where significance has not been attained, the analysis leads to calculation of the timing of the next look, according to the closeness to the stopping barrier. If the calculated timing is higher than 100% information, the trial is stopped immediately for non-significance.*

For each of these sequential methods, the maximum amount of information over and above that of a non-sequential method is remarkably small. Major gains in terms of average information are realized. Maximum sample size requirements for either of the two sequential methods can be derived by multiplying the non-sequential sample size requirement by a relative efficiency factor that depends only upon the Type I and Type II errors.

2. *Power in the General Linear Multivariate Model*

Glueck, Deborah H., Robert Wood Johnson Medical School

Address: Room N-101A University of Medicine and Dentistry of New Jersey Robert Wood Johnson Medical School 675 Hoes Lane Piscataway, NJ 08854-5635

Phone: 908-235-5267

Fax: 908-235-4569

Email: glueckdh@umdnj.edu

Abstract: The general linear multivariate model (GLMM) is used often in clinical and epidemiologic experiments. In contrast to data analysis, power calculations depend on the distributional properties of the predictors. The conditional power is calculated for a specific realization of the predictors. We describe a general class of conditional power approximations based on the method of moments and Taylor series. We demonstrate how existing conditional power approximations fall into this scheme. The unconditional power equals the expected value of the conditional power with respect to all possible stochastic realizations of any random *predictors. The quantile power is the conditional power for a 'worst case' scenario. We summarize unconditional and quantile power results for GLMMs with fixed and Gaussian predictors. An example power analysis illustrates the use of quantile power approximations in study design.*

3. Power for Categorical Data Analysis

O'Brien, Ralph G., Cleveland Clinic Foundation

Address: Director, Collaborative Biostatistics Center Department of Biostatistics and Epidemiology Cleveland Clinic Foundation/P88 9500 Euclid Ave. Cleveland, OH 44195

Phone: 216-445-9451

Fax: 216-444-8023

Email: robrien@bio.ri.ccf.org

Shieh, Gwowen, Cleveland Clinic Foundation

Abstract: From this broad topic we will focus on a pragmatic way to determine power for likelihood ratio tests in applications of logit, log-linear, and Poisson modeling from the simple to the complex. Specifically, we will show via examples how the dominating term of a good general noncentrality approximation for generalized linear models (Self, Mauritsen, and Ohara; Biometrics, 1992) can be computed by forming and analyzing "exemplary" data using one's regular statistical software (as per O'Brien; SUGI-11, 1986). Powers or minimal sample sizes are then computed, tabled, and/or graphed using UnifyPow, a freeware SAS macro (http://www.bio.ri.ccf.org/power.html). The handout will also cover how UnifyPow easily handles statistical planning problems for many other types of categorical data analyses.

4. Sample Size and Power Calculations for Studies with Correlated Data

Liu, Guanghan (Frank), Merck Research Laboratories

Address: Clinical Biostatistics Merck Research Laboratories 10 Sentry Parkway, BL3-2 Blue Bell, PA 19422

Phone: 610-397-2499
Fax: 610-397-2931
Email: guanghan_frank_liu@merck.com

Abstract: Correlated data occur frequently in biomedical research and have received a good deal of attention in recent years. Examples include longitudinal studies, family studies, and opthalmologic studies, etc. Methods to analyze correlated data have been extensively studied in recent years. In this talk, we present methods to compute sample sizes and statistical powers for studies involving correlated observations (Liu and Liang; Biometrics, 1997). This is an important step when the investigators are planning a study to address some scientific hypotheses. The proposed method is based on generalized estimating equations (GEE). The theoretical background and some additional assumptions needed for the sample size and power calculations are highlighted. Sample size and power calculation formulas will be obtained for some special cases that are commonly seen in practice. The methods will be illustrated by simulations and examples from various clinical studies including longitudinal and cross-over trials.

List of speakers who are nonmembers: None

15.2 asa.stat.consult.02

ASA-Statistical Consulting

Session Slot: 4:00- 5:50 Sunday
Estimated Audience Size: 100-125
AudioVisual Request: One Overhead and One Slide Projector

Session Title: Practicing Statistics at the Interface of Science and Policy

Theme Session: Yes
Applied Session: Yes

Session Organizer: **Young, Linda J.** University of Nebraska

Address: Department of Biometry University of Nebraska Lincoln, NE 68583-0712
Phone: 402-472-2903
Fax: 402-472-5179
Email: biom025@unlvm.unl.edu

Session Timing: 110 minutes total (Sorry about format):

> Opening Remarks by Chair - 0 minutes First Speaker - 30 minutes Second Speaker - 30 minutes Third Speaker - 30 minutes Discussant - 15 minutes Floor Discusion - 5 minutes

Session Chair: **Ross, N. Phillip** Director for the Center for Environental Statistics, USEPA

Address: 401 M Street, SW Room 3101 Mail Stop 2163 Washington DC 20450
Phone: 202 260-5244
Fax: 202 260-8550
Email: ross.np@epamail.epa.gov

1. Interface of Science and Policy on Salmon Recovery Efforts in the Columbia River Basin

McDonald, Lyman L., WEST, Inc.

Address: WEST, Inc. 2003 Central Ave. Cheyenne, WY 82001
Phone: 307-634-1756
Fax: 307-637-6981
Email: lymanmcd@csn.net

Abstract: The decline of abundance of salmon in the Columbia River Basin, public policies, and recovery efforts will be briefly reviewed. My view of the role of science in these recovery efforts will be discussed with special emphasis on the interface of statistics, other science, and the public process for making policy and funding decisions.

2. Must We Choose between Production Agriculture and the Environment?

Young, Linda J., University of Nebraska

Address: Department of Biometry University of Nebraska Lincoln, NE68583-0712
Phone: 402-472-2903
Fax: 402-472-5179
Email: biom025@unlvm.unl.edu

Abstract: A brief history of the use of pesticides and fertilizers in production agriculture is presented. From the agricultural perspective, the benefits of these chemical inputs are reviewed. The environmental concerns that have developed from the use of chemicals are discussed. Statistical support for both perspectives and current efforts toward integration of the two approaches are considered.

3. Using Statistics to Determine Data Adequacy for Environmental Policy Decisions under Public Scrutiny

Kelly, Elizabeth J., Los Alamos National Laboratory

Address: Team Leader for Environmental Risk Assessments (TSA-11) Los Alamos National Laboratory Mail Stop K557 Los Alamos, New Mexico
Phone: 505-667-2356
Fax: 505-667-5531
Email: ekelly@lanl.gov
Black, Kelly , Neptune and Co.
Michael, Dan, Neptune and Co.

Abstract: This paper presents EPA's Data Quality Assessment (DQA) approach for data collection and assessment in support of environmental policy decisions. This approach uses a statistical framework, based on the specification of acceptable levels for potential decision errors, to guide data collection and subsequent data assessments. The paper describes two case studies (at

two DOE facilities) where the DQA process was used to support policy decisions about potential risks to the public from exposures to contaminants. These decisions had important political consequences and were subject to intense public scrutiny.

Discussant: **Olsen, Anthony R.** US Environmental Protection Agency

Address: U.S. EPA Environmental Research Laboratory 2385 NW Estaview Circle Corvallis, OR 97330-1067
Phone: 541-754-4790
Fax:
Email: tolson@heart.cor.epa.gov
List of speakers who are nonmembers: TBN

15.3 asa.stat.consult.03

ASA-Statistical Consulting

Session Slot: 10:30-12:20 Wednesday
Estimated Audience Size: 100-125
AudioVisual Request: One Overhead and One Slide Projector

*Session Title: The Client*Consultant Interaction in the Pharmaceutical Industry*

Theme Session: No
Applied Session: Yes

Session Organizer: **Boyd, Michael N.** Pharmaceutical Research Associates, Inc.

Address: Pharmaceutical Research Associates. Inc. 2400 Old Ivy Rd. Charlottesville, VA 22903-4826
Phone: 804-977-2772
Fax: 804-293-7084
Email: mike@pra-ww.com

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 5 minutes First Speaker - 25 minutes Second Speaker - 25 minutes Third Speaker - 25 minutes Discussant - 20 minutes Floor Discusion - 10 minutes

Session Chair: **Boyd, Michael N.** Pharmaceutical Research Associates, Inc.

Address: Pharmaceutical Research Associates, Inc. 2400 Old Ivy Rd. Charlottesville, VA 22903-4826
Phone: 804-977-2772
Fax: 804-293-7084
Email: mike@pra-ww.com

1. Statistical Consulting in a Major Pharmaceutical Firm

Liddle, Roger F., Glaxo Wellcome, Inc.

Address: Director, US Clinical Statistics Glaxo Wellcome, Inc. P.O. Box 13358 5 Moore Dr. Research Triangle Park, NC 27709
Phone: 919-483-5375
Fax: 919-483-6357
Email: liddle-rf@glaxowellcome.com
McSorley, Ellen, Glaxo Wellcome, Inc.

Abstract: The clinical statistician has a critical and well-defined role in the drug development process. Members of the internal drug development team from clinical, regulatory, planning and writing represent one set of clients. Regulatory authorities and the medical community are also clients, and interaction and communication are quite different in this setting. Opportunities to broaden the current role of the clinical statistician will be presented. Comparisons will be made to other opportunities for a consulting statistician.

Another role for the pharmaceutical statistician occurs when statistical analysis is contracted out and the consulting statistician becomes the client. Various models for contracting out statistical work will impact the client/consultant relationship, but communication is the critical success factor in all of them.

A large pharmaceutical firm also holds interesting challenges for the pre-clinical and non-clinical statistician. With problems ranging from research to manufacturing, there are numerous opportunities for the consulting statistician to add value. Examples will be presented.

2. Interaction between the Client and Consultant in the Data Management, Analysis, and Reporting of Clinical Trial Data

Flora, Roger E., Pharmaceutical Research Associates, Inc.

Address: Pharmaceutical Research Associates, Inc. 2400 Old Ivy Rd. Charlottesville, VA 22903-4826
Phone: 804-961-2183
Fax: 804-293-7084
Email: floraroger@pra-ww.com

Abstract: *In the pharmaceutical industry, consultants and consulting organizations are being used more and more to provide the necessary data management and statistical services for accomplishing drug development. Services that may be required by a client cover a broad spectrum, and the specific needs must be communicated. Methods of analysis and the presentation of results for a given trial or drug development program can vary and require decisions to be made at various stages of the analysis and reporting process. Specific requirements in this regard, and the level of decision-making authority to be assigned to the consultant or organization providing the services, must be clearly delineated. Finally, both parties are responsible for certain activities in the process, and careful coordination of these activities is required in order to meet timelines and to provide a well-integrated final product.*

The key to a successful client-consultant interaction in accomplishing the required tasks is good communication throughout the process. This presentation will address steps that can be taken by both the client and the consultant to ensure that the interaction required for the successful completion of a project occurs. Specific experiences will be used to illustrate the ideas presented.

3. An FDA Reviewer's Perspective of the Client-Consultant Interaction in the Pharmaceutical Industry

Kammerman, Lisa A., US Food and Drug Administration

Address: Division of Biometrics II (HFD-715 Center for Drug Evaluation and Research Food and Drug Administration 5600 Fishers Lane Rockville, MD 20857

Phone: 301-827-4257

Fax: 301-827-4264

Email: kammerman@cder.fda.gov

Abstract: *From my perspective as an FDA statistical reviewer, consultants to the pharmaceutical industry are primarily university faculty, independent consultants, and members of CROs. Their familiarity with the regulatory process and requirements is wide-ranging. Consultants provide support in two general areas: (1) the regulatory process from the development of INDs through the submission of NDAs and (2) the development of statistical methodology for the design and analysis of clinical studies.*

For each of these areas, I will discuss the impact of the background of the consultants and their familiarity with the regulatory process on the success, from my perspective, of the client-consultant interaction. A common theme to a successful outcome is the consultant's knowledge of the drug class, indication, and regulatory requirements. Additionally, in the case of an NDA submission, the consultant's knowledge of the database is crucial.

Discussant: **Koch, Gary G.** University of North Carolina-Chapel Hill

Address: Department of Biostatistics McGavran-Greenberg, CB#7400 University of North Carolina Chapel Hill, NC 27599

Phone: 919-966-7282

Fax: 919-966-3804

Email:

List of speakers who are nonmembers: None

Chapter 16

ASA Statistical Education (3)

Program Chair: **Moreno, Jerry L.**

16.1 asa.stat.educ.01

ASA Statistical Education Section

Session Slot: 4:00-5:50 Sunday
Estimated Audience Size: 125-175
AudioVisual Request: Two Overheads

Session Title: Making Statistics Accessible to Students With Disabilities

Theme Session: Yes
Applied Session: Yes

Session Organizer: **Blumberg, Carol Joyce** Winona State University

Address: Dept. of Mathematics and Statistics, Winona State University Winona MN 55987-5838
Phone: 507-457-5589
Fax: 507-457-5376
Email: wncarolj@vax2.winona.msus.edu

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 5 minutes First Speaker - 40 minutes Second Speaker - 40 minutes Floor Discusion - 25 minutes

Session Chair: **Blumberg, Carol Joyce** Winona State University

Address: Dept. of Mathematics and Statistics, Winona State University Winona MN 55987-5838
Phone: 507-457-5589
Fax: 507-457-5376
Email: wncarolj@vax2.winona.msus.edu

1. Access Through Science: Opening the Profession to Students With Disabilities

Preston, Carol, National Oceanic Atmosperic Administration (NOAA)

Address: National Ocean Service, Charleston Laboratory, 219 Fort JJohnson Road, Charleston SC 29412
Phone:
Fax:
Email: carol.preston@noaa.gov

Abstract: Individuals with disabilities are significantly underrepresented in the fields of mathematics, statistics, science, and engineering. There are many reasons for this, but, most significantly, the cause can be attributed to the failure of the educational system, at all levels, to understand and to adequately develop, the capacity of those who are disabled, to undertake science and science-related subjects. Ignorance and misconceptions about the impact of certain disabling conditions, failure to understand the relative ease and low cost of many accommodations, and a failure to capitalize on the increasing wealth and sophisticated capability of technology to minimize physical limitations have meant that many students who are disabled, from their earliest years in school, are actively discouraged from entering into careers that offer challenges and opportunities for personal growth. Several strategies, including mentoring and recruitment, that can lead to optimizing opportunities for disabled students will be presented.

2. Making Statistics Courses More Accessible to Students with Disabilities

Kroeger, Sue Ann, University of Minnesota

Address: Disability Services, University of Minnesota, 16 Johnston Hall, 101 Pleasant St. SE, Minneapolis MN 55455
Phone:
Fax:
Email: sue_kroeger@dsmail.stu.umn.edu

Abstract: This talk will begin with a short discussion of the Americans with Disabilities Act (ADA) and what higher education institutions and instructors are required to do under the law. The talk will then address how instructors, their departments, and their universities can design introductory and other statistics courses to be more accessible to students with disabilities. Both technological and non-technological issues will be discussed. Instructional strategies to implement for students with various physical disabilities (e.g., visually impaired, hearing impaired, wheelchair user) as well as students with a variety of learning disabilities will be given.

List of speakers who are nonmembers: Carol Preston, Sue Ann Kroeger

16.2 asa.stat.educ.02

ASA Statistical Education Section

Session Slot: 8:30-10:20 Wednesday
Estimated Audience Size: 125-175
AudioVisual Request: Two Overheads

Session Title: Plugging In - Connecting Technology to the Classroom

Theme Session: No
Applied Session: Yes

Session Organizer: **Short, Thomas H.** Villanova University

Address: Dept. of Mathematical Sciences, Villanova University, Villanova PA 19085-1699
Phone: 610-519-6961
Fax: 610-519-6928
Email: short@monet.vill.edu

Session Timing: 110 minutes total (Sorry about format):

> Opening Remarks by Chair - 0 minutes First Speaker - 30 minutes Second Speaker - 30 minutes Third Speaker - 30 minutes Discussant - 15 minutes Floor Discusion - 5 minutes

Session Chair: **Short, Thomas H.** Villanova University

Address: Dept. of Mathematical Sciences, Villanova University, Villanova PA 19085-1699
Phone: 610-519-6961
Fax: 610-519-6928
Email: short@monet.vill.edu

1. Dynamic Statistics(TM) - Learner Centered Visualization and Construction of Statistical Concepts

Finzer, William, Key Curriculum Press

Address: Key Curriculum Press, P. O. Box 2304, Berkeley, CA 92702
Phone: 510-548-2304
Fax: 510-548-0755
Email: bfinzer@mail.keypress.com

Abstract: A constructivist outlook on the teaching and learning of statistics suggests that learners should be in charge of devising statistical measures and controlling visualizations of statistical ideas. When students construct their own statistical measures and graphs, they can appreciate the origins of the traditional measures and graphs. Dynamic dragging allows students to change data and observe the effect of the change during the *drag. This paper focuses on visualization techniques, the importance of dragging in Dynamic Statistics(TM) software, and on the role of user-constructed measures in the learning of statistics. It includes results of research from the DataSpace project at Key Curriculum Press, an NSF-funded project to develop a computer learning environment and curriculum materials for use in teaching introductory statistics. Examples from classroom field testing are included.*

2. Virtual Laboratories in Probability and Statistics

Siegrist, Kyle, University of Alabama in Huntsville

Address: Dept. of Mathematical Sciences, University of Alabama in Huntsville, Huntsville AL 35899
Phone: 205-890-6470
Fax: 205-890-6173
Email: siegrist@math.uah.edu

Abstract: A widely accepted principle of the reform movement in mathematics education is that subjects be taught from multiple points of view with appropriate use of technology. This principle is particularly important in probability and statistics. For these subjects, the points of view should include mathematical analysis (the usual derivation of the theory and methods), data analysis (both graphical and numerical, and preferably with real data sets), and simulation analysis. It is unfortunate that the last component is often neglected in standard courses, because carefully designed computer "applets" can be of great value, illustrating the theory and techniques in an interactive and dynamic way that is impossible to achieve by other methods.

The "Virtual Laboratories" project (http://www.math.uah.edu/ stat/) consists of web-based modules that combine hypertext, graphics, interactive Java applets, and data sets. The web is an ideal medium for a project of this type, because the elements can be combined in a carefully designed way, and because the project is available to any student with web access, on any of the standard platforms. Moreover, the project can be linked to a vast array

of related sites and can be continually improved and augmented. The ultimate goal of the project is to have modules that cover the core topics at the undergraduate level, as well as a continually expanding "virtual library" of special random processes.

In this presentation, we will discuss the overall structure of the project, see some examples, and discuss some of the finer design principles, including the special challenges involved in writing hypertext and Java applets.

3. *Simulation and Bootstrapping for Teaching Statistics*

Hesterberg, Tim, MathSoft/Statistical Sciences

Address: MathSoft/Statistical Sciences, 1700 Westlake Ave. N, Suite 500, Seattle QA 98109-3044
Phone: 206-283-8802 x319
Fax: 206-283-0347
Email: timh@statsci.com

Abstract: Some key ideas in statistics and probability are hard for students, including sampling distributions. Computer simulation lets students gain experience with and intuition for these concepts. Bootstrapping can reinforce that learning, and provide a way for students (and future practitioners!) to estimate sampling distributions when they have data but do not know the underlying distribution. Bootstrapping also frees us from the requirement to teach inference only for statistics for which simple formulas are available—we can bootstrap robust statistics like the median as easily as the mean.

For the promise of simulation and bootstrapping to be realized, they must be available and easy to use in general-purpose statistical software, complete with the exploratory data analysis and inferential capabilities required in the course and practice. We discuss some of the available software.

Discussant: **Lock, Robin** St. Lawrence University

Address: Mathematics Department, St. Lawrence University, Canton NY 13617

Phone: 315-229-5960
Fax: 315-229-5804
Email: rlock@vm.stlawu.edu

16.3 asa.stat.educ.03

ASA Statistical Education Section

Session Slot: 10:30-12:20 Tuesday
Estimated Audience Size: 125-175
AudioVisual Request: Two Overheads

Session Title: The Advanced Placement Statistics Exams: The First Two Years

Theme Session: No
Applied Session: Yes

Session Organizer: **Whittinghill, Dexter C.** Rowan University

Address: Dept. of Mathematics, Rowan University, 201 Muillica Hill Rd., Glassboro NJ 08028
Phone: 609-256-4500 x3879
Fax: 609-256-4921
Email: whittinghill@rowan.edu

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 5 minutes First Speaker - 25 minutes Second Speaker - 25 minutes Third Speaker - 25 minutes Discussant - 20 minutes Floor Discusion - 10 minutes

Session Chair: **Whittinghill, Dexter C.** Rowan University

Address: Dept. of Mathematics, Rowan University, 201 Muillica Hill Rd., Glassboro NJ 08028
Phone: 609-256-4500 x3879

Fax: 609-256-4921
Email: whittinghill@rowan.edu

Rosemary A. Roberts, Department of Mathematics, Bowdoin College, Brunswick, Maine. 25 min.

1. *As the Dust Settles: The 1997 and 1998 Advanced Placement Statistics Exams.*

Roberts, Rosemary A., Bowdoin College

Address: Dept. of Mathematics, Bowdoin College, Brunswick ME 04011
Phone: 207-725-3566
Fax: 207-725-3123
Email: rroberts@bowdoin.edu

Abstract: The first Advanced Placement Statistics examination was taken by 7600 students - more than twice the number expected. This paper discusses the types of questions that might appear on an AP Statistics exam, giving examples from the first two exams. A brief description of the AP Statistics Course and exam specifications will be given as background information.

2. *What's a Rubric?: Scoring the Open Ended Questions*

Olsen, Chris, George Washington High School

Address: George Washington High School, Cedar Rapids, IA
Phone:
Fax:
Email: olsen@snert.crwash.k12.ia.us

Abstract: The concept of rubric grading will be discussed, and examples from the AP Statistics tests will be used for illustrative purposes. In addition, the role of "Wholistic" as opposed to "Analytic" grading will be addressed. The construction of a rubric will be modeled, and its evolution traced through the complete AP Statistics test grading process. Since student responses are many and varied, the process is not without some surprises and perplexities; examples of students' surprises will be shared, and sympathy elicited!

Richard L. Scheaffer, Department of Statistics, University of Florida, Gainesville, FL. 25 min.

3. Why did Pat get a 3?: Making the Grade in AP Statistics

Scheaffer, Richard L., University of Florida

Address: Department of Statistics, University of Florida, Gainesville FL 32611-8545
Phone: 904-392-1941
Fax: 904-392-5175
Email: scheaffe@stat.ufl.edu

Abstract: For each AP program, the Reading is an Event. What happens at a Reading of free-response questions will be discussed, along with what happens after the Reading to turn raw scores into the grades received by the students. An interesting part of the process involves a comparability study to see how AP scores relate to college grades in a similar course. Additional discussion will center on how the AP program is being received by colleges around the country, how colleges are handling the scores, and what statisticians might do to help the process.

Discussant: **Velleman, Paul F.** Cornell University

Address: Dept. of Statistics, Cornell University, Ithaca NY 14853-3901
Phone:
Fax:
Email: pfv2@cornell.edu

Chapter 17

ASA Statistical Graphics (3)

Program Chair: **Wegman, Edward J.**

17.1 asa.stat.graph.01

ASA-Stat Graphics

Session Slot: 2:00- 3:50 Monday
Estimated Audience Size: 200
AudioVisual Request: VCR, Slide Projector

Session Title: Real Success Stories of Statistical Graphics

Many papers on statistical graphics tend to focus on doing conventional things with graphics. The real power of graphics is being able to achieve results which could not be achieved with conventional analytic or numerical techniques. The papers in this session focus on describe graphical methods and results in application areas that likely would not have been achieved at all, or would have been achieved with cumbersome, convoluted traditional methods.

Theme Session: Yes
Applied Session: Yes

Session Organizer: **Wegman, Edward J.** Center for Computational Statistics, George Mason University

Address: Center for Computational Statistics George Mason University, MS 4A7 Fairfax, VA 22030-4444
Phone: (703) 993-1691
Fax: (703) 993-1700
Email: ewegman@gmu.edu

Session Timing: 110 minutes total (Sorry about format):

110 minutes total First Speaker - 30 minutes Second Speaker - 30 minutes Third Speaker - 30 minutes Floor Discussion - 10 minutes

Session Chair: **Denby, Lorraine** Bell Labs

Address: Bell Labs, Room 2C255 700 Mountain Avenue, Murray Hill, NJ 07974-0636
Phone: (908) 582-3292
Fax: 908-582-3340
Email: ld@bell-labs.com

1. ***Applications of Statistical Graphics to Discriminant Analysis***

Solka, Jeffrey L., Naval Surface Warfare Center-Dahlgren

Address: COMMANDER ATTN: Dr. Jeffrey L. Solka CODE B10 NAVAL SURFACE WARFARE CENTER 17320 DAHLGREN ROAD DAHLGREN, VA 22448-5100
Phone: 540 653-1982
Fax: 540 653-4930
Email: jsolka@nswc.navy.mil
Marchette, David J., Naval Surface Warfare Center-Dahlgren

Abstract: This talk will explore some of our recent work in the area of discriminant analysis. The specific focus of the talk will be the use of modern statistical graphics procedures to aid in the discriminant analysis process. Examples will be given from a plethora of application areas including chemical detection and classification, identification of regions of interest in unmanned aerial imagery, computer intruder detection, and acoustic signal processing.

2. Counts - Proportions - Interactions: A View on Categorical Data

Theus, Martin, AT&T Laboratories

Address:
Phone:
Fax:
Email:
Wilhelm, Adalbert, University of Augsburg, Germany

Abstract: *Statisticians have developed many graphical displays for continuous data. By contrast, displays for categorical data have been rare. Barcharts can visualize univariate count data, but fail to give a multivariate insight into the data. Although mosaic plots, proposed in 1981 by Hartigan and Kleiner, form a powerful recursive visualization technique for multivariate categorical data, they have not proven popular. The reason for their lack of popularity is the fact that their visual impact depends on the order of the variables. This disadvantage is critical in static implementations of mosaic plots, such as in SAS or S-Plus.*

Embedding mosaic plots in an interactive environment, resolves the limitations of static implementations. Interactive barcharts and mosaic plots offer linked highlighting in multiple views. Interaction also facilitates to rearrange the order of variables as well as the order of levels.

Mosaic plots can be augmented to modeling-tools for categorical data by superimposing residual information. Direct interaction allows the user to perform graphical stepwise modeling. Linked highlighting opens the door to visualize categorical response models. This represents a new methodology that goes beyond traditional exploratory uses of data visualization.

3. Large Scale Genome Analysis Approaches: Integrating Bioinformatics and Data Mining

Somogyi, Roland, NINDS/NIH

Address: Molecular Physiology of CNS Development LNP/NINDS/NIH Bldg.36/Rm.2C02 Bethesda, MD 20892
Phone: 301-402-1407
Fax: 301-402-1565
Email: rolands@helix.nih.gov
Michaels, George, Institute for Computational Sciences and Informatics, George Mason University

Abstract: Making the transition from finding DNA sequences, to identifying coordinated functions, requires an integrated approach to the study of gene expression. A combined approach using quick surveys via a variety of methods and quantitative kinetic analysis of gene expression is being implemented by several groups. A comparison of different analysis and visualization approaches of these data to identify coherent patterns of expression and control in time series data will be presented.

List of speakers who are nonmembers: Roland Somogyi, George Michaels. Adalbert Wilhelm, Martin Theus

17.2 asa.stat.graph.02

ASA- Statistical Graphics

Session Slot: 10:30-12:20 Tuesday
Estimated Audience Size: 200
AudioVisual Request: VCR, slide projector

Session Title: Statistical Graphics on the Web

The world wide web is arguably positioned to re-revolutionize statistical computation in much the same way that individual workstations and personal computers did during the 1980s. JAVA and ActivX offer the possibility of doing statistical graphics over the web in a way that cannot necessarily be easily duplicated on individual workstations. Special types of graphics not available in standard packages can be offered over the web. This session explores aspects of this advanced mechanism for distributing the latest graphics.

Theme Session: Yes
Applied Session: No

Session Organizer: **Wegman, Edward J.** Center for Computational Statistics, George Mason University

Address: Center for Computational Statistics George Mason University, MS 4A7 Fairfax, VA 22030-4444
Phone: (703) 993-1691
Fax: (703) 993-1700
Email: ewegman@gmu.edu

Session Timing: 110 minutes total (Sorry about format):

110 minutes total First Speaker - 30 minutes Second Speaker - 30 minutes Third Speaker - 30 minutes Floor Discussion - 15 minutes)

Session Chair: **Oster, Robert A.** Morehouse School of Medicine

Address: Clinical Research Center Morehouse School of Medicine 720 Westview Drive, S.W. Atlanta, GA 30310-1495
Phone: 404-756-5713
Fax: 404-752-1112
Email: BOster@compuserve.com

1. Web-Based Visualization

Eick, Stephen G., Bell Labs/Lucent Technologies

Address: Bell Labs/Lucent Technologies - Rm 1G-351
1000 East Warrenville Road
Naperville, IL 60566

Phone: 630-713-5169
Fax: 630-713-4982
Email: eick@research.bell-labs.com

Abstract: *Over the last couple years we have developed a suite of web-based interactive systems using Java applets. These include systems for visualizing software source code data, displaying time-stamped network traffic, performing interactive data analysis, and an interface for controlling network simulations. Our experience has been that Java is a more productive programming environment than C++ and has the potential to become a standard development platform for building visualization systems.*

2. Untangling the Web: Java Beans for Disseminating Information Through Interactive Graphics

Rope, Daniel J., Daniel J. Rope Consulting

Address: 1870 North Scott Street #548, Rosslyn, VA 22209-1353
Phone: 703/525-6404
Fax:
Email: dan_rope@mnsinc.com
Carr, Daniel B., George Mason University
Wilkinson, Leland, SPSS, Inc.

Abstract: *The Internet, intranets, and extranets have all proven to be effective for allowing vast numbers of people access to vast amounts of information. But, just like other communications media before them, these networks create new challenges in the process solving previous problems. In particular, the new challenges include creation of quality presentations and customized interaction with data from databases.*

To address these issues, we will describe and demonstrate a new set of statistical graphics Java Beans designed specifically for web-based data dissemination. This new system focuses on three aspects of statistical graphics: powerful content creation, customizable interactivity with rich meta-data, and flexible, high-quality presentation. The content creation engine provides a simple way to model complicated data structures and provide flexible mapping of data values to the aesthetic attributes of graph elements. In addition, customizable "controller" beans give content providers complete control over end user interactivity. Further, *a rich meta-data model allows content providers to embed important contextual information into a graph. This information can be used in a hyper-linked environment and can also be used to warn end users when they attempt to make inappropriate comparisons. Finally, a variety of carefully designed templates for high quality are embedded in customizable beans. These allow refined control of the look and feel of graphs. These three aspects: a content engine, customizable interactivity with a meta-data model, and carefully designed templates provide an integrated solution to the majority of challenges in web-based data dissemination.*

3. The Data Image: A Tool for Exploring High Dimensional Data Sets

Minnotte, Michael C., Utah State University

Address: Utah State University, Department of Mathematics and Statistics Logan, UT 84322-3900
Phone: (801) 797-1844
Fax: 801/797-1822
Email: minnotte@SUNFS.math.usu.edu
West, R. Webster, University of South Carolina

Abstract: The color histogram was first introduced as a tool for visualizing higher dimensional data by Wegman (1990). A new version of this concept, called a data image, is discussed. Each variable is transformed into a greyscale or color range so that a high-dimensional data set may be viewed as an image, with observations on one axis and variables on the other. The rows and columns of the image may be rearranged to highlight relationships between variables. New ways of displaying the image based on various linear orderings of a data set are discussed. The ability of the data image to show relationships between variables and clusters in higher dimensions is explored. A Java applet which may be used to construct an interactive data image via the World Wide Web is used as the basis for this exploration.

List of speakers who are nonmembers: None

17.3 asa.stat.graph.03

ASA Statistical Graphics

Session Slot: 10:30-12:20 Thursday
Estimated Audience Size: 200
AudioVisual Request: VCR, slide projector

Session Title: Graphics for Data Mining

Data mining as both a technology and a buzz word has become the darling of the computer science and business applications world. At first sight, this would seem to be nothing more than a repackaging of well-known techniques of exploratory data analysis. However, requirements of data set size and complexity often make traditional exploratory graphical techniques unusable. New approaches which can simultaneously deal with massive data sets, high dimensionality and missing data are requirements of new graphical techniques. This session explores some of these directions.

Theme Session: Yes
Applied Session: No

Session Organizer: **Wegman, Edward J.** George Mason University

Address: Center for Computational Statistics George Mason University, MS 4A7 Fairfax, VA 22030-4444
Phone: 703-993-1691
Fax: 703-993-1700
Email: ewegman@gmu.edu

Session Timing: 110 minutes total (Sorry about format):

> 110 minutes total First Speaker - 30 minutes Second Speaker - 30 minutes Third Speaker - 30 minutes Discussant - 10 minutes (or none) Floor Discussion - 10 minutes

Session Chair: **Symanzik, Juergan** George Mason University

Address: Center for Computational Statistics MS 4A7 George Mason University 4400 University Drive Fairfax, VA 22030-4444
Phone: (703) 993-3786
Fax: (703) 993-1700
Email: symanzik@galaxy.gmu.edu

1. A Successful System Using Data Mining for Direct Marketing Response Prediction

Wendelberger, James G., Consultant

Address: 620 Los Pueblos Los Alamos, NM 87544-2615
Phone: 505 662 2838
Fax: 505 662 2838
Email: jimwen@sisna.com

Abstract: The "GAIN" system to accomplish data mining for marketing response data is described. The system was utilized in the "Knowledge and Data Discovery Cup '97 Competition" to tie for first place. The system is utilized for direct marketing response prediction. Practical aspects of the creation of the system are described. Problems which have been solved are presented. Research areas are described.

2. Visualising Large Datasets: Issues and Ideas

Stone, Glenn, CSIRO Mathematical and Information Sciences, Australia

Address: CSIRO Mathematical and Information Sciences, Locked Bag 17, North Ryde, NSW 2113, Australia
Phone: +61 2 9325 3100
Fax: +61 2 9325 3200
Email: Glenn.Stone@cmis.csiro.au

Lovell, David, CSIRO Mathematical and Information Sciences, Australia

Abstract: In the typical application areas of data mining, for example finance and insurance, datasets can consist of millions of *observations and hundreds of records. This has significant implications for* traditional *visualisation methods such as scatter plots and scatter plot matrices, coplots and 3D rotating point clouds. We consider techniques applicable to the data mining problems of exploratory data analysis, market segmentation and identifying unusual groups.*

3. Crystal Vision: A Graphical Tool for Data Mining

Luo, Qiang, George Mason University

Address: Center for Computational Statistics George Mason University, MS 4A7 4400 University Drive, Fairfax, VA 22030-4444
Phone: (703) 993-3648
Fax: (703) 993-1700
Email: qluo@jupiter.galaxy.gmu.edu

Abstract: Crystal Vision is an extension of the SGI based software known as ExplorN, the latter developed as a research demonstration tool at George Mason University. The intent is to have a high performance software operating on a UNIX workstation with the capability of real-time or near real-time graphical analysis of large, high dimensional data sets. The software is capable of handling 20 to 30 dimensional data and it has been tested with 250,000 observations. This talk will discuss both the computational considerations taking advantage of some of the architectural features of the Silicon Graphics workstation as well as some of the statistical and graphical considerations.

Discussant: **Poston, Wendy L.** Office of Naval Research

Address: Office of Naval Research Division of Mathematical and Computer Sciences 800 N. Quincy Street, Arlington, VA 22217-5000
Phone: (703) 696-4320
Fax: (703) 696-2611
Email: postonw@onr.navy.mil
List of speakers who are nonmembers: None

Chapter 18

ASA Statistics and the Environment (3 + 2 competition)

Program Chair: **Bunck, Christine M.**

18.1 asa.stat.environ.01

ASA-ENVR

Session Slot: 8:30-10:20 Monday
Estimated Audience Size: 50-60
AudioVisual Request: overhead & slide projector only

Session Title: Design of Environmental Monitoring Networks

Theme Session: Yes
Applied Session: Yes

Session Organizer: **Royle, J. Andy** National Center for Atmospheric Research

Address: Geophysical Statistics Project, NCAR, Boulder, CO 80307
Phone: 303-497-1704

Fax: 303-497-1333
Email: rolye@ncar.edu

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 0 minutes First Speaker - 35 minutes Second Speaker - 35 minutes Third Speaker - 35 minutes Floor Discusion - 5 minutes

Session Chair: **Royle, J. Andy** National Center for Atmospheric Research

Address: Geophysical Statistics Project, NCAR, Boulder, CO 80307
Phone: 303-497-1704
Fax: 303-497-1333
Email: royle@ncar.edu

1. Shared Network Redesign Decision Making

Haas, Timothy C., University of Wisconsin-Milwaukee

Address: School of Business Administration, University of Wisconsin-Milwaukee, P.O. Box 742, Milwaukee, WI 53201
Phone: 414-229-4360
Fax: 414-229-6957
Email: haas@uwm.edu

Abstract: Generally, if the public feels they have a say in determining how environmental monitoring will be conducted, they will be more willing to support long-term monitoring expenditures. Hence, publicly-accessible systems are needed that allow anyone to participate in monitoring network design and redesign decision making. This talk will describe how such a system can be implemented as a World-Wide-Web site.

The proposed system will allow a web site user to perform the same site addition and site deletion optimization analyses that are used by governmental decision makers. The focus of this talk will be on the optimal removal of sites from a monitoring network.

It will be argued that (a) network optimality is inherently multi-dimensional and hence requires multiple objectives to be satisfied, and (b) the public needs to be involved in the determination of the these objectives and their priority weights. It is also argued that there are three core objectives of any monitoring network. First and second, the network should deliver pollutant predictions that are as accurate and precise as possible. Third, the network should provide these predictions as inexpensively as possible.

For a network monitoring a large random field, stationarity is typically not present. Also, the problem of identifying the best subset of sites for deletion is combinatorially explosive. For these two reasons, Simulated Annealing is used to solve the nonlinear, integer, multi-objective mathematical programming problem using the author's Moving Cylinder Kriging spatio-temporal prediction algorithm to compute average prediction differences between the full and reduced network (network accuracy), and average increase in prediction relative error between the full and reduced network (network precision).

An example using NADP/NTN sulfate deposition across the conterminous U.S will be given.

2. Monitoring Network Design and Air Quality Standards

Sampson, Paul D., University of Washington

Address: Department of Statistics, Box 354322, University of Washington, Seattle, WA 98195-4322

Phone: 206-685-2664

Fax: 206-685-7419

Email: pds@stat.washington.edu

Guttorp, Peter, University of Washington

Thompson, Mary Lou, University of Washington

Cox, Larry, U.S. Environmental Protection Agency

Abstract: A number of issues are generally taken into account in designing an air quality monitoring network. Local siting criteria for individual monitoring sites are well-discussed in the literature, but formal statistical design criteria seem not to be much *used. We discuss issues of spatial and temporal sampling and averaging in the definition of air quality standards, the consequent specification of violators of these standards, and the probability of detecting violators from the perspective of monitoring network design for spatio-temporal air quality processes.*

3. Space-time Models and Dynamic Design

Wikle, Christopher K., National Center for Atmospheric Research

Address: NCAR, Boulder, CO 80307
Phone: 303-497-1722
Fax: 303-497-1333
Email: wikle@ucar.edu
Royle, J. Andy, National Center for Atmospheric Research

Abstract: An important area of spatial statistics is the design of spatial sampling plans for monitoring air pollution and other environmental processes. Many methods for constructing optimal spatial designs are widely applied in a large number of disciplines. These methods generally produce static designs which are optimal under models with no explicit temporal structure. However, most environmental processes often tend to exhibit both spatial and temporal variability, and hence static networks may not capture the essential space-time variability of the process. Although static designs are often necessary due to geopolitical and economic considerations, one could imagine the possibility of supplementing a static network with one or more mobile monitoring devices. Then, the design problem is where should the mobile monitor or monitors be located at time t+1 in order to provide the most information based on observations up to time t. We consider several approaches to this problem, including a simplistic approach which moves the mobile monitor to the location which is optimal at time t. We then propose a dynamical space-time model that allows prediction of future spatial correlation structure, and use this correlation structure to optimize over all possible monitoring sites. These approaches are demonstrated with a surface ozone data set.

List of speakers who are nonmembers: 0

18.2 asa.stat.environ.02

ASA-ENVR

also sponsored by the Statistical Ecology Section of the Ecological Society of America

Session Slot: 2:00- 3:50 Wednesday
Estimated Audience Size: 40-50
AudioVisual Request: overhead projector

Session Title: Hierarchical Models in Ecology

Theme Session: No
Applied Session: No

Session Organizer: **Edwards, Don** University of South Carolina

Address:
Phone:
Fax:
Email: edwards@math.sc.edu

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 0 minutes First Speaker - 30 minutes Second Speaker - 30 minutes Third Speaker - 30 minutes Discussion lead by chair - 10 minutes Floor Discussion - 10 minutes

Session Chair: **Dixon, Philip** Savannah River Ecology Laboratory

Address: University of Georgia, Savannah River Ecology Lab, P.O. Box E, Aiken, SC 29802-0005
Phone: 803-725-2472
Fax: 803-725-3309
Email: dixon@srel.edu

1. *Applications of Conditionally Specified Models in Ecology*

Kaiser, Mark S., Iowa State University

Address: Department of Statistics, Iowa State University, Snedecor Hall, Ames, IA 50011-1210
Phone: 515-294-8871
Fax: 515-294-4040
Email: mskaiser@iastate.edu
Cressie, Noel, Iowa State University
Lee, J., Iowa State University
Lewin, N., Iowa State University
Daniels, M., Iowa State University

Abstract: *Conditionally specified statistical models show promise for applications involving random variables that exhibit complex dependence structures. Such models are formulated by specifying a Markov random field on a lattice, the 'locations' of which may be physical (i.e., spatial lattices) or relational (i.e., longitudinal). A probability density or mass function is specified for each of the locations, conditional on all other locations in the field. The goal of statistical analysis is then to identify the joint distribution and use it for estimation and inference about pertinent parameters. In this talk, conditionally specified models will be applied to several problems in ecology. At least one of these applications will involve dependence arising from spatial sources, and at least one will involve dependence arising from multiple observations being taken on the same sampling units. Dependence in the models may be located at either the level of the data model or in a mixing distribution for random data model parameters.*

2. *Hierarchical Models for Analyzing Spatial Ecological Data*

Ver Hoef, Jay M., Alaska Department of Fish and Game

Address: Alaska Department of Fish and Game, 1300 College Road, Fairbanks, Alaska 99701
Phone: 907-459-7278
Fax: 907-452-6410

Email: ffjmv@aurora.alaska.edu

Abstract: *Consider a spatial lattice of locations where data have been observed, and the data are assumed to be a realization of a hierarchical model. The hidden process is assumed to have a (prior) distribution from a two-state (e.g., high or low) Markov model. The two states form patches in the hidden process. However, the patches are hidden by random noise. Conditional on the states, the observations may be modeled as coming from any distribution in the regular exponential class. An objective of some ecological studies is to uncover the patches in the hidden process, and make some further inference, such as computing a landscape metric, on the patch structure. Sampling from the posterior distribution of the hidden process allows inference on the patch structure and the mean and variance of any landscape metric computed from the sample. In this paper I consider one and two dimensional lattices and use parametric empirical Bayes and fully hierarchical Bayes models to sample from the posterior distribution of the hidden process and make inference on metrics from landscape ecology.*

3. Mixed-Model Splines for Ecological Time Series

Street, IV, W. Scott, Georgia Southern University

Address: Department of Math. And Computer Science, Georgia Southern University, P.O. Box 8093, Statesboro, GA 30460-8093
Phone: 912-681-5888
Fax: 912-681-0654
Email: wsstreet@gasou.edu
Edwards, Don, University of South Carolina

Abstract: *Ten-year biweekly time series of abundance of benthic fauna in a South Carolina salt marsh reveal, in the log scale, a very clear annual pattern of alternating near-linear growth and decline. The ecological processes governing these dynamics, particularly the dramatic year-to-year differences in annual maxima*

and minima, are mysterious. Traditional time series analyses (Fourier and ARIMA methods) can (in some cases) adequately mimic the patterns in these series, but they do little to enhance ecological understanding because they are not meaningfully parameterized. We model these series using floating-knot linear splines in two ways: (1) as high-dimensional nonlinear regression models with no forced similarity from year to year; and (2) as a hierarchical spline model placing a probability distribution on annual maxima, minima, growth rate, and decay rate across years, thereby enforcing some similarity in seasonal patterns. We discuss fitting and inference under both approaches; both yield excellent fits with ecologically interpretable results which can be used to shed light on governing processes. Results of a simulation study show that the hierarchical approach can yield great improvements in efficiency over the nonlinear regression approach, particularly for series with low interannual variability.

List of speakers who are nonmembers: 0

18.3 asa.stat.environ.03

ASA-ENVR

Session Slot: 8:30-10:20 Thursday
Estimated Audience Size: 40-50
AudioVisual Request: overhead & slide projector

Session Title: Statistical Multiscale Assessment of Landscapes and Watersheds with Satellite and Synoptic Data

Theme Session: No
Applied Session: No

Session Organizer: **Patil, G.P.** Pennsylvania State University

Address: Center for Statistical Ecology and Environmental Statistics Department of Statistics The Pennsylvania State University 421-D Thomas Bldg. University Park, PA 16802

Phone: 814-865-9442
Fax: 814-865-1278
Email: gpp@stat.psu.edu

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 5 minutes First Speaker - 25 minutes Second Speaker - 25 minutes Third Speaker - 25 minutes Fourth Speaker - 25 minutes Floor Discussion - 5 minutes

Session Chair: **Patil, G.P.** Pennsylvania State University

Address: Center for Statistical Ecology and Environmental Statistics Department of Statistics The Pennsylvania State University 421-D Thomas Bldg. University Park, PA 16802
Phone: 814-865-9442
Fax: 814-865-1278
Email: gpp@stat.psu.edu

1. Quantitative Characterization of Hierarchical Scaled Landscape Patterns

Taillie, C., Pennsylvania State University

Address: Center for Statistical Ecology and Environmental Statistics Department of Statistics The Pennsylvania State University 421-C Thomas Bldg. University Park, PA 16802
Phone: 814-865-5212
Fax: 814-865-1278
Email: taillie@stat.psu.edu
Patil, G.P., Pennsylvania State University

Abstract: When a natural landscape is cast as a categorical raster map, a multiresolution characterization of spatial pattern can be obtained whereby the entropy is computed for a finer resolution map, conditioned on the values of a coarser resolution map. After application to a sequence of rescaled maps which have increasingly degraded resolution, the conditional entropy is plotted as a function of measurement scale (resolution), thus resulting in *a multiresolution profile of fragmentation patterns. For neutral landscapes that are simulated by multiresolution stochastic generating models, we present a method to directly compute conditional entropy profiles. Such profiles can provide benchmarks for comparing results obtained from raster maps of actual landscapes that are classified from satellite images. Results show that characteristic landscape types give rise to characteristic features of these fragmentation (conditional entropy) profiles.*

2. Guided Transect Sampling for Assessing Sparse Populations

Ringvall, Anna, Swedish University of Agricultural Sciences

Address: Department of Forest Resource Management Swedish University of Agricultural Sciences S-901 83 Umea, SWEDEN
Phone: 46-90-7866838
Fax: 46-90-778116
Email: anna.ringvall@resgeom.slu.se
Stahl, Goran, Swedish University of Agricultural Sciences

Abstract: *In guided transect sampling, prior information is used to guide the field survey within randomly laid out broad strips. Generally, the prior information consists of remote sensing data. The strips are divided into grid-cells, and a covariate value is estimated for each such cell. The covariate data could be, e.g., estimates of the biomass of hardwoods in case the intention is to survey some population which prefers such stand and site conditions. Different strategies can be used for guiding the surveyor through the strips, based on the covariate values in the grid-cells. The different strategies lead to different properties of the estimators of population size, etc. In this paper, the general principles of guided transect sampling are outlined, and results from theoretical and practical tests of the method are reported.*

3. Spatial and Spectral Classification of Compressed Image Data for Landscape Analysis

Filipponi, Danila, University of Chieti, Italy

Address: until 6/1/98 Visiting Faculity Center for Statistical Ecology and Environmental Statistics Department of Statistics The Pennsylvania State University 421-B Thomas Bldg. University Park, PA 16802; after 6/1/98 Department of Quantitatiave Methods and Economic Theory Universita degli Studi "G. D'Annunzio", Viale Pindaro, 42, 65127 Pescara, ITALY

Phone: 814-863-8126

Fax: 814-865-1287

Email: danila@stat.psu.edu

Abstract: *Remote sensing data are becoming increasingly important in environmental studies. Two limitations have been the high cost of acquiring these proprietary data sets and the disk storage and processing demands imposed by their large sizes. Data compression is one approach for addressing these limitations. Here, pixels are clustered into a fixed number of categories on the basis of their spectral response values. Quantitative response information is partially retained by recording a set of summary statistics for each cluster.*

Central issues are the development of methodology for analyzing the compressed data. The paper examines these issues in the specific context of supervised classification into eight landcover types and both spectral and spatial considerations enter into the final classification. A spectral classification using the Kullback-Liebler distance between clusters is proposed and is compared with a more ad hoc approach using only the information in the diagonals of the summarizing variance-covariance matrices. Spatial coherence of the resulting landcover map has been assessed by an iterative method involving the use of indicator kriging.

4. Wavelet-based Multiscale Analysis in Landscape Ecology

Li, Bai-Lian, University of New Mexico

Address: Department of Biology University of New Mexico Albuquerque, NM 87131-1091

Phone: 505-277-5140

Fax: 505-277-0304

Email: blli@unm.edu

Abstract: *Wavelet analysis is relatively a new mathematical theory and computational method. Because of its characteristics of time/space-frequency localization and multiresolution, the wavelet transform of a signal can provide detailed information of underlying ecological processes in time and/or spatial scale. In wavelet representation, a signal is decomposed into a sum of elementary building blocks describing its local frequency content. This paper describes the use of wavelet analysis as an innovative hierarchical data analysis technique for investigating multiscale spatio-temporal relationships in several long-term ecological research data sets. We use wavelet transforms, wavelet variance, cross-scale correlation analysis of transform coefficients, wavelet-based change point detection, and wavelet-based local self similarity index to characterize and quantify heterogeneous landscape patterns on a wide range of temporal and spatial scales. Preliminary results demonstrate that these methods are suitable to characterize multiscale spatio-temporal data and the scaling problem. The scaling information and the proposed approach provide a new way to look at the long-term dynamics of the coupling between the terrestrial ecosystem, the hydrological cycle, and the physical climatic system, and has potential to give a detailed extrapolation across temporal or spatial scales.*

List of speakers who are nonmembers: 3, Goran Stahl (replaced by Anna Ringvall 3-4), Danila Filipponi, and Bai-lian Li

18.4 asa.stat.environ.04

Submitted by ASA-ENVR for competition

Session Slot: Sunday, 4:00-5:50
Estimated Audience Size: 50-60
AudioVisual Request: overhead and slide projectors

Panel Session Title: *How Low Can We Measure: The Statistics of Detection and Quantitation*

Theme Session: Yes
Applied Session: Yes

Session Organizer: **Coleman, David** Alcoa

Address: Applied Math & Computer Tech. (483), ATC-D-10 Alcoa Technical Center, 100 Technical Drive, Alcoa Center, PA 15069
Phone: 724-337-5913
Fax:
Email: david.coleman@alcoa.com

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 5 minutes Panelists Comments/Discussion - 85 minutes Floor Discussion - 20 minutes

Abstract: *A multitude of statistical issues surrounds the seemingly simple question: "How low a concentration or level can we reliably measure or detect?" This question is pertinent to environmental sciences and analytical chemistry, but also to medical sciences and national defense. In this session, four distinguished panelists with diverse background present and tackle some of these issues, then discuss and debate the merits of various solutions. One concern is how to balance detection sensitivity with false positives/false alarms. Also important is selecting a population subsample; estimating between-lab and within-lab sources of bias and variation; developing calibration methods that will work in the face of heteroscedasticity and the presence of outliers; describing the relation between measurement uncertainty and significant digits; and applying equivalence testing.*

The first panelist will comment on the statistical and conceptual properties of various estimators for the quantification limit and address issues of bias, variability and robustness. The second panelist will describe the calculation of detection limits for ion-chromatographic data and discuss the strength and weaknesses of two approaches. The third panelist will comment on the application of statistical models to accommodate heteroscedasticity and

traditional laboratory approaches for estimating detection limits. The fourth panelist will describe the newly adopted ASTM standard practice for detection and the proposed standard practice for quantification.

This session addresses the theme of statistics as a guide for policy. Everyone has or should have an interest in how low a concentration - such as lead in water, PCB's in soil, carbon monoxide in air - can reliably be measured. The requirement to protect human health and ecosystems has caused regulatory agencies to push for lower limits on pollutants, but the ideal 0% is unattainable; so policy must be established to determine acceptably low limits. That policy should be based on statistics, analytical chemistry, and practical considerations.

This session will be of interest to applied statisticians. Because all measurements are uncertain, statistics and analytical chemistry must be used jointly to estimate the lowest concentration that can be reliably detected and the lowest concentration at which one can reliably report a numerical measurement value. Issues at hand that will be discussed and are of interest to applied statisticians are selection of sample population, biases (e.g., between labs), heteroscedasticity, blind trials, false detections, power of detection, significant digits, and calibration.

This session will be of interest to members of other ASA sections including the Sections on Physical & Engineering Sciences, Government Statistics, Biometrics, and Health Policy. Statisticians, environmental scientists, and policy-makers in government regulatory agencies, USGS, DoD, and NIST will be interested in this topic.

Session Chair: **Wendelberger, Joanne R.** Los Alamos National Laboratory

Address: Los Alamos, NM 87545
Phone: 505-665-4840
Fax: 505-667-4470
Email: joanne@lanl.gov
Gibbons, Robert University of Illinois

Address: Chicago, IL

Phone: 312-413-7755
Fax: 312-996-2113
Email: robert.gibbons@uic.edu
Vanatta, Lynn Air Liquide

Address: Dallas, TX
Phone: 972-995-7541
Fax:
Email: ALAV%mimi@magic.itg.ti.com
Davis, Charles Environmetrics & Statistics Limited

Address: Envirostat, Environmetrics and Statistics Ltd., 1853 Wellington,Henderson, NV 89014
Phone: 702-456-8994
Fax: 702-456-8114
Email: bromero@compuserve.com
Coleman, David Alcoa

Address: Applied Math & Computer Tech. (483), ATC-D-10 Alcoa Technical Center, 100 Technical Drive, Alcoa Center, PA 15069
Phone: 724-337-5913
Fax: 724-337-4911
Email: david.coleman@alcoa.com
List of speakers who are nonmembers: 1 - Lynn Vanatta

18.5 asa.stat.environ.05

Submitted by ASA-ENVR for competition; Co-sponsored by the International Association for Mathematical Geology

Session Slot: 8:30-10:20 Tuesday
Estimated Audience Size: 75-100
AudioVisual Request: xxx

Session Title: Advances in Geostatistics

Theme Session: No
Applied Session: No

Sponsored by: ASA Section on Statistics and the Environment (ASA-ENVR) and by the International Association for Mathematical Geology (IAMG). The primary aim of the IAMG is to promote international cooperation in the application and use of mathematics in geological research and technology. In keeping with this aim, IAMG is interested in fostering increased collaboration with statistical organizations, hoping that this will result in increased cross-disciplinary research. Two of IAMG's journals, Mathematical Geology and Computers and Geosciences, regularly publish state-of the-art theory and methodology pertaining to environmental statistics, spatial statistics, and geostatistics. Many ASA/ASA-ENVR statisticians, particularly those whose research interests include spatial statistics, are members of IAMG or read its journals.

Session Content: The purpose of the session is to familiarize the JSM audience with the statistical work of internationally-renown researchers in other disciplines that routinely use and develop statistical methodology, most notably in geostatistics, in their research. The intent is to foster future collaboration/educational opportunities between interested JSM participants and IAMG scientists and to encourage ASA participation at international conferences sponsored by the IAMG or other, similar international organizations. Most of the speakers for this session do not usually attend the JSM, and many JSM members may not attend the international meetings that often serve as the primary forum for the introduction of new methods in spatial statistics and geostatistics. Since these speakers do not normally attend the JSM, necessary funding for these speakers will be provided by the IAMG.

Session Organizer: **Gotway Crawford, Carol A.** Centers for Disease Control and Prevention

Address: Centers for Disease Control and Prevention, Atlanta, GA 30333
Phone: (770) 488-7428
Fax: (770) 488-7335
Email: cdg7@cdc.gov

Session Timing: 110 minutes total (Sorry about format):

110 minutes total First Speaker - 30 minutes Second Speaker - 30 minutes Third Speaker - 30 minutes Discussant - 15 minutes Floor Discussion - 5 minutes

Session Chair: **Gotway Crawford, Carol A.** Centers for Disease Control and Prevention

Address: Centers for Disease Control and Prevention, Atlanta, GA 30333
Phone: (770) 488-7428
Fax: (770) 488-7335
Email: cdg7@cdc.gov

1. *Geostatistical Estimation Applied to Highly Skewed Data*

Clark, Isobel, Geostokos Limited

Address: Geostokos Limited, Alloa Business Centre, Whins Road, Alloa FK10 3SA Central Scotland
Phone: +44 1259 211904
Fax: +44 1259 211904
Email: isobel@stokos.demon.co.uk

Abstract: Many variables in the geological and earth sciences follow highly skewed distributions. For example: precious metal grades; toxic pollution of groundwater aquifers; grain size distributions and so on. Transformations can be applied to 'normalise' such data and geostatistical studies have been undertaken on these transforms. The major problem is to 'backtransform' predicted values to provide unbiased and reliable estimates for the original measurements. This presentation will use a case study on lognormal data to illustrate various backtransformations which have been published and/or used in practice. In particular, the problem of predicting the average value over an area or volume will be considered in detail.

2. *Conditional Simulation of Random Sets*

Lantuéjoul, Christian, Ecole des Mines de Paris

Address: Ecole des Mines de Paris, Centre de Géostatistique, 35 rue Saint-Honoré, 77305 Fontainebleau, France
Phone: (33) (1) 64 69 47 60
Fax: (33) (1) 64 69 47 05
Email: lantu@cg.ensmp.fr

Abstract: *Predicting the geometry of a porous medium is a standard problem encountered in many domains of applications. It is often tackled by simulations (Monte Carlo procedure) which provide not only a prediction, but also its variability.*

Any simulation exercise requires a stochastic model describing the statistical properties of the porous medium. This model can be defined either explicitly or implicitly, and in the second case, it is not always easy to check whether it makes sense. This question will be considered first.

In order to reduce the variability of the prediction, it is preferable to consider only conditional simulations. These are simulations compatible with the available information, typically honoring data points, satisfying domain proportions or respecting connectivity constraints. The second part of the presentation will be devoted to a review of conditional simulation algorithms as a function of the model considered and of the set of conditions to be respected. Several examples will be given to illustrate these ideas.

3. Incorporating Model Uncertainty in Geostatistical Methods of Risk Analysis

Dowd, Peter A., Department of Mining and Mineral Engineering

Address: Department of Mining and Mineral Engineering, University of Leeds, Leeds LS2 9JT England, UK
Phone: (44) (113) 233-2796
Fax: (44) (113) 246-7310
Email: P.A.Dowd@leeds.ac.uk

Abstract: *A major use of conditional simulation is as a means of risk assessment. In many earth science applications accurate field measurement of a single value of a variable (e.g. fractures, permeability) is difficult and measurement of all values is impossible. Thus a stochastic approach, such as conditional simulation, is the only effective way of modelling these variables. However, such approaches are based on models which can only be inferred from sparse data. Often the uncertainty on the model is greater than the uncertainty in the variable. The talk will look*

at ways in which model uncertainty can be incorporated into the simulation. In particular, the maximum likelihood approach will be discussed.

Discussant: **Hohn, Michael Ed.** West Virginia Geological and Economic Survey

Address: West Virginia Geological and Economic Survey, P.O. Box 879, Morgantown, WV 26507-0979

Phone: 304-594-2331

Fax: 304-594-2575

Email: hohn@geosrv.wvnet.edu

List of speakers who are nonmembers: Clark, Lantuejoul, Dowd — (Hohn is ASA member 4-4)

Chapter 19

ASA Statistics and Sports (2)

Program Chair: **Albert, James H.**

19.1 asa.stat.sports.01

ASA-SIS

Session Slot: 8:30-10:20 Wednesday
Estimated Audience Size: 100
AudioVisual Request: xxx

Session Title: Measuring Dcompetitive Ability: Current Trends in Rating Systems for Games and Sports

Theme Session: No
Applied Session: Yes

Session Organizer: **Glickman, Mark** Boston University

Address: Department of Mathematics Boston University 111 Cummington St. Boston, MA 02215
Phone: 617-353-5209
Fax: 617-353-8100
Email: mb@math.bu.edu

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 0 minutes First Speaker - 30 minutes Second Speaker - 30 minutes Third Speaker - 30 minutes Discussant - 10 minutes Floor Discussion - 10 minutes

Session Chair: **Bradlow, Eric** University of Pennsylvania

Address: Marketing Department Wharton School of Business University of Pennsylvania 3620 Locust Walk Suite 1400 Steinberg-Dietrich Hall Philadelphia, PA 19104
Phone: 215-573-8255
Fax:
Email: ebradlow@wharton.upenn.edu

1. The ATP Tour Tennis Ranking System: Moving Towards the 21st Century

Dore, Philippe, ATP Tour

Address: ATP Tour 200 ATP Tour Boulevard Ponte Vedra Beach Florida 32082
Phone: 904-285-8000
Fax:
Email: pdore@atptour.usa.com

Abstract: *Since its inception in 1973, the ATP computer ranking system has been the standard by which male professional tennis players have been judged. The ATP Ranking has been used as the basis for acceptance of entries and seeding for most major international professional tennis tournaments. Prior to 1990, a player's total points were divided by the number of events he played with a minimum divisor of 12 tournaments. Since 1990, the ATP Tour has used the Best 14-system which ranks a player based on the total points he earns from his best 14 events. The ATP Tour is currently evaluating new systems, such as Next 6- and 3+3-system.*

2. The United States Chess Federation Rating System: Current Issues and Recent Developments

Jones, Albyn, Reed College

Address: Mathematics Department Reed College 3203 SE Woodstock Blvd Portland, OR 97202
Phone: 503-771-1112
Fax:
Email: jones@reed.edu
Glickman, Mark, Boston University

Abstract: The United States Chess Federation (USCF) rating system, originally developed by Elo, is essentially a simple linear filter. Versions of this system are used by a variety of other organizations, including the World Chess Federation (FIDE) and the US Table Tennis Association. Over time it has become clear that the system does not account for important sources of variability associated with identifiable characteristics of competitive chess players, such as the greater variability of weak players' abilities, and the potential for rapid improvement of scholastic players. This talk discusses the concerns with the earlier system, and the details of the recent overhaul designed to address these issues.

3. The Scrabble Ratings System: A History, the Current State, and a Possible Future

Parker, Robert, University of New Mexico

Address: P.O. Box 3365 Albuquerque, NM 87190-3365
Phone: 505-661-6560
Fax:
Email: rparker@unm.edu

Abstract: The rating system used by the NSA was developed in the early 1980's mainly as a mirror of the chess ratings system. In this talk we discuss some of the differences between the chess and Scrabble populations and the differences in the ratings systems. Performance of the system through 1997 will be displayed and discussed. Also: Post-analysis of 1992-1996 data

using some recently suggested methods, and some recommendations of the 1997-1998 NSA Ratings Committee that sprung from these analyses.

Discussant: **Glickman, Mark** Boston University

Address: Department of Mathematics Boston University 111 Cummington St. Boston, MA 02215
Phone: 617-353-5209
Fax: 617-353-8100
Email: mb@math.bu.edu
List of speakers who are nonmembers: None

19.2 asa.stat.sports.02

ASA-SIS

Session Slot: 2:00- 3:50 Monday
Estimated Audience Size: 100
AudioVisual Request: xxx

Session Title: Rounding the Bases: Modeling Behavior in the Business of Baseball

Theme Session: No
Applied Session: Yes

Session Organizer: **Albert, Jim** Bowling Green State University

Address: Department of Mathematics and Statistics Bowling Green State University Bowling Green, OH 43403
Phone: 419-372-7456
Fax: 419-372-6092
Email: albert@math.bgsu.edu

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 0 minutes First Speaker - 30 minutes Second Speaker - 30 minutes Third Speaker - 30 minutes Discussant - 10 minutes Floor Discussion - 10 minutes

Session Chair: **Albert, Jim** Bowling Green State University

Address: Department of Mathematics and Statistics Bowling Green State University Bowling Green, OH 43403
Phone: 419-372-7456
Fax: 419-372-6092
Email: albert@math.bgsu.edu

1. Can You Tell the Players Without a Scorecard? Errors in Variables in Baseball Discrimination

Wobbekind, Richard L., University of Colorado

Address: Campus Box 419 College of Business University of Colorado Boulder, CO 80309
Phone: 303-492-1147
Fax: 303-492-5962
Email: wobbekin@colorado.edu
Frederick, David M., University of Colorado
Kaempfer, William H., University of Colorado

Abstract: Jackie Robinson's signing by the Brooklyn Dodgers in 1947 ushered a new era for Major League Baseball, signaling the sport's color barrier had been broken. In the subsequent fifty years, baseball has seemed to some a triumph of equal opportunity, where participants are evaluated solely on their merits. To others, questions linger, such as: why do the racial composition of teams and management differ, and does racial discrimination exist in salary or career longevity? Thus it is not surprising that research on discrimination in baseball has expanded. What is unexpected is that in these studies, little attempt was made to consider the intricacies of the concept, race. For example, most of these studies regress the dependent variable (e.g., salary) on a list of productivity indicators and a dummy variable for race.

This dichotomization creates difficulties in estimating the extent of discrimination in baseball. The current project has three goals. The first is to develop a more formal approach for recognizing race and ethnic differences among players that will be helpful to the study of discrimination in baseball. The significant error that could result from dichotomizing a race variable suggests the need for more formality and understanding. The second goal is to use the approach to examine a collective bargaining institution, the baseball salary arbitration mechanism, for any evidence it may yield on discrimination. Last, the approach will be used to revisit studies of discrimination in baseball that use statistical evidence.

2. *Streak Management*

Fort, Rodney, Washington State University

Address: Department of Economics Washington State University Pullman, WA 99164-4741
Phone: 509-335-1538
Fax: 509-335-4362
Email: fort@mail.wsu.edu

Abstract: The paper is about winning streaks in baseball. It analyzes game streaks, within a season, over a couple of seasons. We look for non-random streaks and find very little evidence of any. However, in the frequency domain, an interesting test shows that it certainly does "pay" if one can manage for streaks! Further, fans love them, as evidenced by a model of attendance at major league baseball games

3. *Managerial Succession and Labor Market Response*

Butler, Stephen A., The University of Oklahoma

Address: College of Business Administration The University of Oklahoma 307 West Brooks, Room 200 Norman, OK 73019-0450
Phone: 405-325-5759
Fax:
Email: sbutler@ou.edu

Abstract: *Managers in every organization contract to provide a certain amount of productive effort to the organization. It is often assumed that without some monitor of effort, the manager will shirk and become ineffective. Fama (1980) proposed that both the internal and external labor market for managers' effort would discipline the mangers so that they provide the contracted amount of productive effort. Managerial succession within an organization has concentrated on the effect of that organization. The contention is that a more effective manager is brought in as a replacement. However, little is known about the managerial labor market's response to the change. Is the market able to recognize and appropriately judge the former manger's effectiveness? This study indicates that in a relatively noiseless market the response is appropriate. There are also reputation effects rewarded by the labor market.*

Discussant: **Berry, Scott** Texas A&M

Address: Statistics Department Texas A&M University College Station TX 77843-3143.

Phone: 409-845-5650

Fax: 409-845-3144

Email: berry@stat.tamu.edu

List of speakers who are nonmembers: Wobbekind, Frederick, Kaempfer, Fort, Butler

Chapter 20

ASA Survey Research Methods (5)

Program Chair: **Martin, Elizabeth A.**

20.1 asa.survey.rm.01

SRMS

Session Slot: 8:30-10:20 Wednesday
Estimated Audience Size: xx-xxx
AudioVisual Request: none dws

Session Title: Analysis of Longitudinal Data from Complex Surveys

Often survey data involve repeated measurements from the same individual. Rotating Panel Data provides a good example. In such cases, the assumption of independence for the same individual over different time points is highly questionable. Building dependence among the observations through subject-specific random effects is a standard tool for analyzing such data. The speakers will discuss various aspects of it both frequentist and Bayesian, and will point out a large number of applications for complex surveys.

Theme Session: No
Applied Session: ?

Session Organizer: **Ghosh, Malay** University of Florida

Address:
Phone:
Fax:
Email: ghoshm@stat.ufl.edu

Session Timing: 110 minutes total (Sorry about format):

110 minutes total...please allocate Opening Remarks by Chair - 5 minutes First Speaker - 25 minutes Second Speaker - 25 minutes Third Speaker - 25 minutes Discussant - 20 minutes Floor Discussion - 10 minutes

Session Chair: **Ghosh, Malay** University of Florida

Address:
Phone:
Fax:
Email:

1. Analysis of Longitudinal Data from Complex Surveys

Rao, J. N. K., Carleton University

Address:
Phone:
Fax:
Email:

Abstract: *Semi-parametric marginal generalized linear models are often used for the analysis of longitudinal data. Using such models, Liang and Zeger (1986) obtained standard errors of regression parameter estimates and associated Wald tests, assuming a "working" correlation structure for the repeated measurements on a sample subject. Rotnitzky and Jewell (1990) developed quasi-score tests and Rao Scott adjustments to "working" quasi-score tests. These methods are asymptotically robust to misspecification of the correlation structure, but assume independence of sample* *subjects which is not satisfied for complex survey data based on stratified multistage sampling. We propose asymptotically valid Wald and quasi-score tests for longitudinal survey data, using the jackknife or the Taylor linearization methods. Alternative tests, based on Rao-Scott adjustments to naive tests that ignore survey design features and on Bonferroni-t, are also developed. These tests are particularly useful when the effective degrees of freedom, usually taken as the total number of sample primary units (or clusters) minus the number of strata, is small.*

2. ***Modelling of Complex Survey Longitudinal Data***

Nathan, Gad, Hebrew University of Jeruselum

Address:
Phone:
Fax:
Email:
Pfefferman, Danny, Hebrew University of Jerusalem

Abstract: Longitudinal data consist of repeated measurements on the same sampling units such as households or individuals. Measurements associated with the same unit are naturally intercorrelated, and these correlations need to be accounted for in any model fitting process. Two other important considerations when fitting models to longitudinal data obtained from complex surveys are (1) the causes and pattern of the missing observations, whether implied by the sampling design such as in rotating sample surveys, from leaving the population, or from nonresponse, and (2) the possible effects of the sampling mechanism and in particular, the sample selection probabilities on the model holding for the sample data.

In this paper we study plausible models and inference methods that are suitable for fitting short term intercorrelated data that are subjected to sampling effects and (possibly non-ignorable) missing data. In particular, we explore methods of estimating the intercorrelations and incorporating them in subsequent analysis, with special emphasis on multilevel models. These models are in common use for analyzing data with strong intracluster correlations *that result from common unobservable factors affecting units in the same cluster. The methods will be illustrated using simulated and real survey data obtained from the Israel Labor Force Survey.*

3. Analysis of Longitudinal Binary Data from Complex Surveys: A Hierarchical Bayesian Approach

Lahiri, Partha, University of Nebraska-Lincoln

Address:
Phone:
Fax:
Email: plahiri@unlinfo.unl.edu
Maiti, T., Wu, C., University of Nebraska-Lincoln

Abstract: Longitudinal data analysis has received considerable importance in recent years among survey samplers. In this article, we develop a fully hierarchical model-based methodology to analyse longitudinal binary data that can arise in a complex survey. We carry out our analysis using the Monte Carlo Markov Chain (MCMC) technique. We extend different Bayesian model diagnostic tools to finite population sampling and evaluate them through simulation.

Discussant: **Binder, David** Statistics Canada

Address:
Phone:
Fax:
Email:
List of speakers who are nonmembers: None

20.2 asa.survey.rm.02

SRMS

Session Slot: 10:30-12:20 Tuesday

Estimated Audience Size: 90
AudioVisual Request: xxx

Session Title: The Application of Logistic Models with Mixed Effects to Sample Survey Datasets

In recent years, there has been tremendous progress in developing the algorithms and software to fit models of this type. At least two distinct methodologies have evolved, one frequentist, the other Bayesian. The frequentist approach most favored is the PQL (penalized quasi-likelihood) of Breslow and Clayton. This method has been incorporated into two packages for analysis of multi-level data, HLM and MLn. The Bayesian approach uses Monte Carlo Markov Chain methods such as Gibbs sampling, data augmentation, and the Metropolis-Hastings algorithm. Both methods are very computationally intensive, but the Bayesian methods are more so than the frequentist methods. The disadvantage of the frequentist methods is that they tend to underestimate the variance of predicted random effects. In this session, there will be speakers from both schools. The idea will be for each to present a brief description of the methodology (preferably, accessible to a wider audience), to contrast their methodology with other available methodology, and to present any evaluation results that are available. The most common applications of these methods to sample survey datasets involve small area estimation and imputation. Given the tradition of design-based inference of datasets for public policy purposes, a critical question for the proponents of these new methodologies is how to present their estimates to journalists and politicians. The discussant was chosen specifically to focus on this issue from the randomization-based perspective.

Theme Session: No
Applied Session: Yes

Session Organizer: **Judkins, David R.** Westat

Address:
Phone:
Fax:
Email: jundkind1@westat.com

Session Timing: 110 minutes total (Sorry about format):

110 minutes total...please allocate Opening Remarks by Chair - 5 minutes First Speaker - 25 minutes Second Speaker - 25 minutes Third Speaker - 25 minutes Discussant - 15 minutes Floor Discussion - 15 minutes

Session Chair: **Judkins, David R.** Westat

Address:
Phone:
Fax:
Email:

1. Bayesian Model Averaging for Small Area Estimation with Binary Data

Ghosh, Malay, University of Florida

Address:
Phone:
Fax:
Email: ghoshm@stat.ufl.edu
Natarajan, Kannan, Bristol-Myers Squibb PRI

Abstract: *Hierararchical Bayesian methods are now being widely used for the analysis of generalized mixed linear models (GLMM). One of the advantages of this approach is that it accounts for the uncertainty in the variance components by modeling rather than simple point estimation. These methods have proved to be very useful for small area estimation where one needs estimates simultaneously for several local areas. Sample sizes being small for the individual areas, one needs to "borrow strength" from similar neighboring areas.*

Often there are situations when there is no clearcut choice among several models. In such situations, one can find the posterior probabilities of the different models, pick the one with the highest posterior probability, and find small area estimates and standard errors based on that model. Another option is not to report estimates and standard errors based on a single model, but *report estimates which are weighted averages of estimates based on different contemplated models, the respective weights being proportional to the poterior probabilities of these models.*

We will illustrate the Bayesian model averaging idea with respect to a dataset pertaining to job satisfaction. There are 84 local areas classified according to seven geographic regions and 12 demographic categories. The responses are binary. The analysis is provided by using several hierarchical Bayes GLMM's, and then using the model averaging idea.

2. Model-Based Approaches to Small Area Estimation with Binary Data

McCulloch, Charles E., Cornell University

Address:
Phone:
Fax:
Email: cem1@cornell.edu

Abstract: Small area estimation in sample surveys can be addressed using mixed models for binary data. Methods and software for model-based, frequentist analysis are much more prevalent than a decade ago. I will illustrate the use of mixed models and best predicted values for small area estimation and compare and contrast the various methods available for fitting such models. Several readily-available methods perform poorly for binary data and should be avoided.

3. Estimation by the Method of Generalized Zero Functions for Mixed Nonlinear Models

Singh, Avi, Statistics Canada

Address:
Phone:
Fax:
Email: singhavi@statcan.ca
Wu., S., Statistics Canada

Abstract: *There has been considerable development in recent years in the area of Bayesian and Frequentist approaches to estimation of parameters for mixed nonlinear models. Bayesian methods do provide optimal estimates for all parameters (fixed and random) but the assumptions about the priors are hard to verify. Frequentist methods, on the other hand, are based on much less restrictive distributional assumptions (e.g., up to second moments in a semiparametric framework) but may not give rise to consistent estimates of fixed (first and second order) parameters. In this paper we propose a frequentist solution termed method of generalized zero functions(GZF) which does give consistent estimates of all fixed parameters for a fairly general class of mixed nonlinear models, and optimal estimates of first order parameters (fixed and random). The GZF method can be viewed as a natural generalization of the familiar maximum quasi-likelihood method (or more generally the method of estimating functions) for fixed nonlinear models. This is similar to the way the best linear unbiased prediction for mixed linear models generalizes the method of linear zero functions of C.R. Rao (this is equivalent to BLUE) for fixed linear models. For survey data the GZF method adapts easily the Godambe-Thompson approach for optimal (jointly under design-model) estimation of superpopulation parameters. Illustrative examples for estimation with survey and nonsurvey data will be presented.*

Discussant: **Folsom, Ralph** Research Triangle Institute

Address:
Phone:
Fax:
Email:
List of speakers who are nonmembers: None

20.3 asa.survey.rm.03

SRMS

Session Slot: 2:00- 3:50 Sunday
Estimated Audience Size: 50
AudioVisual Request: none dws

Session Title: Network and Adaptive Sampling–Theory and Application

Conventional sample designs uniquely link only one sampling unit to each population element. With the development in the 1970's of network sampling and the concept of a counting rule, survey researchers at the design stage had available a choice of counting rules that allowed more than one sampling unit to report a population element. Network sampling has been shown to be useful for reducing: (1) sampling errors for estimating population quantities for rare and elusive populations, (2) measurement biases for coverage deficiencies and reporting errors, and (3) dependence in dual system estimation. Network sampling has been applied to many difficult problems such as estimating the prevalence of illicit drug use, diabetes, and cancer, and the undercount in the US Census. Another type of sampling useful for surveying rare populations is adaptive cluster sampling. In adaptive cluster sampling, an initial probability selection of units is made and then after the sampled units with the attribute of interest are identified additional units in a neighborhood of these units are sampled. The neighborhood often consists of sampling units that are geographically close to each other. This type of sampling can be quite efficient when there is geographical clustering of the population of interest. For example, in ecological applications where there is interest in estimating the population size of rare species of plants or animals which tend to grow or live in close proximty to each other.

There are interesting connections in the application and theory of network and adaptive cluster sampling which can be explored. The purpose of this session is bring together practitioners and methodologists working on problems using both types of sampling so that they can present applications and methodological developments.

Theme Session: Yes/No
Applied Session: Yes

Session Organizer: **Graubard, Barry** National Cancer Institute

Address:

Phone:
Fax:
Email: graubarb@dcpcepn.nci.nih.gov

Session Timing: 110 minutes total (Sorry about format):

110 minutes total...please allocate Opening Remarks by Chair - 0 minutes First Speaker - 25 minutes Second Speaker - 25 minutes Third Speaker - 25 minutes Fourth Speaker - 25 minutes Floor Discussion - 10 minutes

Session Chair: **Graubard, Barry** National Cancer Institute

Address:
Phone:
Fax:
Email:

1. The History of Network Sampling

Sirken, Monroe, National Center for Health Statistics

Address:
Phone:
Fax:
Email:

Abstract: This paper reviews the history of network sampling, and the applications of network sampling in circumstances where multiple selection units are naturally linked to the same population elements, and other circumstances where multiple linkages are deliberately fostered to improve survey design efficiency. Also, the paper describes the network sampling paradigm, and the survey counting rule and other important design features of network surveys.

2. Population Based Network Sample Surveys of Establishments

Shimizu, Iris, National Center for Health Statistics

Address: 6525 Belcrest Road, Hyattsville, MD 20782
Phone:
Fax:
Email:
Sirken, Monroe, National Center for Health Statistics, Westat
Judkins, David R., National Center for Health Statistics, Westat

Abstract: *In a population based network sample survey of medical providers (vis. hospitals, physicians, dentists, clinics, etc.), the sample providers are selected from among those reported as being seen by respondents in a household sample survey. The network sample design is suggested whenever good medical provider sampling frames are unavailable. This paper discusses other conditions favoring the network design as well as other factors that may contraindicate its use.*

3. Adaptive Sampling in Graphs
Thompson, Steven K., Pennsylvania State University

Address:
Phone:
Fax:
Email: skt@stat.psu.edu

Abstract: *Adaptive sampling designs are those in which the procedure for selecting the units to include in the sample may depend on values of variables of interest observed during the survey. For example, neighboring units may be added to the sample whenever high values are observed. In spatial sampling the neighborhood is defined by geographic proximity. In studies of human populations the neighborhood may also be defined by social relationships.*

In studies of hidden and hard-to-reach human populations such as injection drug users and others at risk for HIV transmission, adaptive link-tracing designs in which initial respondents lead investigators through social links to other individuals often provide

the only practical way to obtain a sample large enough for the study. Data summaries or inference from such samples can be misleading, however, if the sample-selection procedure is not taken into account. The situation is conceptualized as sampling in a graph, with the nodes of the graph representing people and the arcs or arrows represeting social relationships. The problem is that data are observed for only a sample of the nodes and arcs, from which we wish to infer characteristics of the whole graph or population.

Examples of link-tracing designs include network sampling, snowball sampling, chain-referral methods, "random walk" designs, and adaptive cluster sampling. Design-based and model-based methods of inference with such designs will be discussed in this talk.

4. Using a Restricted Adaptive Sampling to Estimate Fish Larval Abundance

Lo, Nancy C.H., Southwest Fisheries Science Center

Address: P.O. Box 271, La Jolla CA 92038
Phone:
Fax:
Email: NLO@UCSD.EDU
Griffith, David, Southwest Fisheries Science Center
Hunter, John R., Southwest Fisheries Science Center

Abstract: Adaptive sampling is a sampling design in which the procedure for selecting sample sites and allocating sampling effort depends on data collected during the survey. In March 9 - 27, 1995, a restricted stratified adaptive sampling was used to survey Pacific hake larvae in California water because the spatial distribution of Pacific hake larvae is highly patchy. The survey was conducted between Los Angeles and San Francisco covering an area of 202,115 km2 (59,540 nm2). Because of limited survey time, we used a restricted adaptive sampling design imposing a maximum number of station in each stratum. A stratified two-stage cluster Horvitz-Thompson (HT) and a simple stratified

(SS) sample mean were used to estimate mean catch per tow. For our survey design, although HT is biased , the variance of the HT estimate was high. The variance of SS was lower than the variance of sample mean from both simple random sampling and proportional stratified sampling. Simulation confirmed our conclusions. It also estimated bias of SS, and showed that the sample variances of HT and SS were biased downward. In some cases, SRS performed well. Nonetheless, our adaptive sampling was relatively easy to implement, and it provided biological information within patches, which may not have been possible if conventional sampling designs were used.

List of speakers who are nonmembers: None

20.4 asa.survey.rm.04

SRMS

Session Slot: 8:30-10:20 Thursday
Estimated Audience Size: 100
AudioVisual Request: xxx

Session Title: The Controversy about Sampling in the Census

Plans for the 2000 Census have been complicated by controversies about the Census Bureau's proposed use of sampling in the census. This session presents different perspectives in the debate about sampling, in order to shed light on the issues and contribute to the resolution of differing points of view.

Theme Session: Yes
Applied Session: Yes

Session Organizer: **Obstat, Neal**

Address:
Phone:
Fax:

Email:

Session Timing: 110 minutes total (Sorry about format):

110 minutes total. Opening Remarks by Chair - 5 minutes First Speaker - 25 minutes Second Speaker - 25 minutes Third Speaker - 25 minutes Closing Remarks by the chair - 15 minutes Floor Discussion - 15 minutes

Session Chair: **Brown, Lawrence** Wharton

Address:
Phone:
Fax:
Email: lbrown@stat.wharton

1. The Case Against Sampling and Estimation, Or Why Accuracy Does Matter

Bensen, Clark, Polidata

Address: 3112 Cave Crt., Lake Ridge, Virginia
Phone:
Fax:
Email: polidata@aol.com

Abstract: The speaker will discuss the intent of the Framers at the Constitutional Convention of 1787 with respect to the federal census and its role in safeguarding an objective and accurate standard for both representation and taxation. He will also talk about the critical role the census plays in the political redistricting process and why accuracy should be considered from the perspective of low-level census geography.

2. Census 2000: The Statistical Case

Killion, Ruth Ann, U. S. Bureau of the Census

Address:
Phone:

Fax:
Email:

Abstract: *This paper reviews the work that led to the statistical decisions for Census 2000. The decisions cover all aspects of sample design and estimation for the One Number Census the bureau plans to conduct in 2000. We include census methodology from precensus through mailout, sampling for census nonresponders, conducting the Integrated Coverage Measurement Survey, processing, estimation and postcensus activities. We discuss some of the options considered and describe the process used to choose the final methods.*

The paper also addresses some of the issues the Census Bureau faces as it tries to educate and convince people about the soundness of our plans. We have learned that technical words (for example, sampling), while concise and descriptive to the knowledgeable, have totally different meanings colloquially. And those informal understandings are often counter to the technical meanings.

3. Who Counts? Census Controversies and the Millenial Census

Anderson, Margo, University of Wisconsin-Milwaukee

Address: Department of History, University of Wisconsin–Milwaukee, WI 53201
Phone:
Fax:
Email: margo@csd.uwm.edu
Fienberg, Stephen E., Carnegie-Mellon University

Abstract: *The presentation will review the major points of our forthcoming book, Who Counts?, which analyzes the legal, political and statistical controversies surrounding of the 1990 census. We will relate these controversies to the policy questions framing the 2000 count, and discuss the relative responsibilities of Congress, the Census Bureau, the Department of Commerce,*

and the statistical community in shaping the methods to be used in 2000. We hope to provide some guidance to the key actors in the 2000 plans, and disentangle the complex legal, political, demographic and statistical issues involved in taking the modern census.

List of speakers who are nonmembers: Bensen

20.5 asa.survey.rm.05

SRMS

Session Slot: 2:00- 3:50 Monday
Estimated Audience Size: 100
AudioVisual Request: xxx

Session Title: Achieving Quality in Surveys

Theme Session: No
Applied Session: Yes

Session Organizer: **Lyberg, Lars** Statistics Sweden

Address:
Phone:
Fax:
Email: lars.lyberg@scb.se

Session Timing: 110 minutes total (Sorry about format):

110 minutes total...please allocate Opening Remarks by Chair - 10 minutes First Speaker - 20 minutes Second Speaker - 20 minutes Third Speaker - 20 minutes Floor Discussion - 20minutes

Session Chair: **Biemer, Paul P.** Research Triangle Institute

Address:
Phone:

Fax:
Email:

1. ***Quality Improvement in Surveys - A Process Perspective***
Biemer, Paul P., Research Triangle Institute
Lyberg, Lars, Statistics Sweden

Address:
Phone:
Fax:
Email: lars.lyberg@scb.se

Abstract: Survey quality is directly related to survey errors. As for sampling errors comprehensive theories exist. For nonsampling errors no such theory exists. Left uncontrolled, nonsampling errors can render the resulting survey data useless for many important survey objectives. Typically specific error sources are treated one by one by using control or verification procedures and sometimes there are integrated treatments or simultaneous modeling of several specific error sources. Most of these attempts lead to estimates of post-survey quality indicators which are important for deciding data accuracy, but, except for repeated surveys, may be of little value for improving the survey data. Therefore, interest must shift from post-survey quality evaluation to controlling the survey processes such as data collection, data processing and data analysis. Process quality generates product quality.

Many survey organizations throughout the world are now working with the concepts of TQM in the context of survey design and survey execution. Methods for process quality such as control charting, quality teams, customer focus and decision-making based on scientific methods are now being successfully applied in survey work. It is our intention to provide examples of such attempts.

2. ***Reporting on Data Quality and Process Quality***
Dobbs, Joy, Office for National Statistics, U.K.

Address:

Phone:
Fax:
Email:
Davies, Pam, Office for National Statistics, U.K.
Martin, Jean, Office for National Statistics, U.K.
Ruddock, Vera, Office for National Statistics, U.K.

Abstract: *Users of statistics and survey data need to have information which enables them to assess the quality of the data collected: they need to be able to judge whether the data are suitable for the purposes envisaged. In Britain the Official Statistics Code of Practice asserts that the Government Statistical Service (GSS) has a duty to "provide guidance and interpretation to help users understand and use the statistics." It goes on to indicate the sort of information that should be provided to aid such interpretation. To help put this into practice, the GSS recently published a Statistical Quality Checklist which consists of a detailed list of questions which should be considered when describing the statistics or survey data in a report or publication.*

Underlying the checklist is the recognition that, although ideally users require information about data quality, in practice there are many aspects of data quality which cannot be measured - or only with considerable effort and expense. It is therefore important also to provide information about the quality of the processes used in collecting and producing the statistics: process quality is often a good indicator of data quality when direct measures of the latter are not available.

This paper discusses the sort of information that can be produced to measure data and process quality, drawing on the statistical approach to quality measurement exemplified by the concept of total survey error with its emphasis on identifying and measuring errors from different sources. However, this approach is set within a broader definition of quality derived from a TQM perspective. Here the emphasis is on customer focus and the provision of a complete service to meet customer needs, providing the right information, in the right form, at the right time and price, and which goes beyond the production of high quality data.

3. Measuring Survey Quality in a CASIC Environment

Couper, Mick P., Survey Research Center, University of Michigan and Joint Program in Survey Methodology

Address:
Phone:
Fax:
Email:

Abstract: *Computer assisted survey information collection (CASIC) is changing the way the survey data collection process is designed, implemented and evaluated. Methods like CATI and CAPI have made quality control or evaluation tools used in paper and pencil interviewing (e.g., review of paper forms, item-missing data counts, skip errors) obsolete. At the same time, field staff (interviewers, supervisors, managers, trainers, etc) are being asked to take on new and more complex tasks. CASIC thus offers both new challenges for the measurement and evaluation of the survey process, as well as providing the opportunity to develop new tools and procedures for such evaluation. This paper will review some of the key impacts of automation on quality measures and discuss alternative approaches and procedures for developing quality measures in a CASIC environment. These include both design and testing tools (e.g., usability testing) as well as tools to evaluate the data collection process (e.g., automated production statistics, keystroke file analysis).*

4. Improving Survey Quality Through Pretesting

DeMaio, Theresa, U. S. Bureau of the Census

Address:
Phone:
Fax:
Email:

Abstract: *As the vehicle of data collection, the questionnaire is one of the critical components in achieving high quality in a*

survey. The best of sampling schemes and estimation strategies will not yield accurate data if the answers provided by the respondent are not meaningful.

In recent years, there has been increased emphasis on building quality into the questionnaire design process through pretesting. Methods have been developed or adopted from other fields that focus on the response process, using either the respondent alone or the respondent's interaction with the interviewer as the target of study. Laboratories have been established in the U.S. statistical agencies to implement the methods. And generic clearance procedures that foster pretesting of government questionnaires have been established.

In this paper, we describe three methods that are used at the U.S. Census Bureau to pretest questionnaires: cognitive interviewing, behavior coding of respondent/interviewer interaction, and respondent debriefing. Examples of their use in testing U.S. Census Bureau survey questionnaires will be presented, and examples of their results will be documented.

List of speakers who are nonmembers: None

Chapter 21

ASA Teaching of Statistics in the Health Sciences (1)

Program Chair: **Freeman, Daniel**

21.1 asa.teach.stat.hs.01

ASA Teaching of Statistics in the Health Sciences

Session Slot: 2:00- 3:50 Tuesday
Estimated Audience Size: 100-200
AudioVisual Request: xxx

Session Title: Should We Only Teach Them What They Need To Know To Pass The Boards?

Theme Session: Yes/No
Applied Session: Yes

Session Organizer: **Ackerson, Lynn** Kaiser Permanente Medical Care Program

Address: 3505 Broadway, Oakland, CA 94611
Phone: 510-450-2386
Fax: 510-450-2071

Email: LMA@dor.kaiser.org

Session Timing: 110 minutes total (Sorry about format):

Invited Session Opening Remarks by Chair - 5 or 0 minutes First Speaker - 30 minutes (or 25) Second Speaker - 30 minutes Third Speaker - 30 minutes Discussant - none Floor Discusion - 20 minutes

Session Chair: **Ackerson, Lynn** Kaiser Permanente Medical Care Program

Address: 3505 Broadway, Oakland, CA 94611
Phone: 510-450-2386
Fax: 510-450-2071
Email: LMA@dor.kaiser.org

1. What They Test and Why

Dawson, Beth, Southern Illinois University

Address: Dept. of Internal Medicine Southern Illinois University School of Medicine PO Box 19230 Springfield, IL 62794-9230
Phone: 217-524-2441
Fax: 217-524-8156
Email: bdawson@siumed.edu

Abstract: *"It's on the boards" - a thinly disguised threat many of us have used to motivate our students to learn and apply basic statistical concepts and critical appraisal skills. Yet, many of us who teach would be hard-pressed to say exactly what content is, indeed, "on the boards." And, if we knew, would we agree that the concepts being tested are the important ones that physicians and other health professionals need to know? We might be interested in the criteria that certifying boards and testing agencies use to determine the topics they test.*

This presentation focuses on two issues regarding the examinations written and administered by National Board of Medical Examiners and those examinations written and administered *by several of the larger specialty certification boards in medicine. First will be a description of the content and concepts contained on the tests and the cognitive level at which they are tested. Second will be a summary of the procedures and resources medical specialty boards use to determine and validate the statistical content on their examinations.*

2. The Use of Targeted Lectures and Small Groups For Teaching Statistics to Medical Students

Freeman, Daniel H., University of Texas Medical Branch

Address: Department of Preventive Medicine and Community Health University of Texas Medical Branch Galveston TX

Phone: 409 772-6616

Fax: 409 772-9127

Email: dan.freeman@utmb.edu

Philips, Billy U., University of Texas Medical Branch

Hokanson, James A., University of Texas Medical Branch

Abstract: *Statisticians, as with most professionals, teach medical students the material that statisticians believe to be useful in the practice of medicine. Standard texts cover descriptive statistics, probability, and inference. This and a brief introduction to epidemiology and critical appraisal of the medical literature constituted the second year course in Epidemiology and Biostatistics at UTMB up until 1993. We were confronted with both declining student interest and declining test scores, when we began to restructure the course. A careful review of the guidelines for the USMLEs revealed that while these topics were covered, the emphasis needed to be changed. Specifically, we found a greater emphasis on test characteristics, basic survival analysis, graphics and study design than had been previously expected. In addition, we found the students wanted an opportunity to interact directly with the faculty. In response, the content of the lectures was revised to reflect a greater emphasis on clinical issues and a closer alignment to USMLE topics. We also began small groups sessions of 2 faculty and 12 students to review papers from the literature.*

We have not been able to document a specific improvement in test scores, but there has been a marked improvement in student satisfaction. As an unanticipated benefit, faculty have reported a general satisfaction with the program and an improvment in the non-clinical faculty's perception of the medical students.

3. 10,000 Monkeys Can Get a Perfect Score On the National Boards – Can You?

Murphy, James R., University of Colorado Health Sciences Center

Address: Dept. of Preventive Medicine and Biometrics Box B-119 4200 E. 9th Ave. Denver, Co 80262
Phone: 303-315-5170
Fax: 303-315-3183
Email: James.Murphy@uchsc.edu

Abstract: *Teaching statistics to medical students has never been an enjoyable occupation. Periodically new ideas such as problem based learning, evidence based medicne, medical decision analysis, ROC analysis, etc. provide new techniques and motivating examples, but it continues to be difficult to interest students in the underlying statistical principles. At the same time, national reports such as Physicians for the Year 2000, indicate that clinicians need to understand these principles in order to deal with the exponential growth of medical research. As statisticians involved in this process, should we be teaching skills, attitude, process, zeitgeist, and/or the enjoyment of learning? The answer is easy – all of the above. How best to do this remains to be decided. This talk will cover some of my most and least successful interactions with medical students. I hope to provide some ideas about how to teach all things to all people in a way that is beneficial to teacher and student alike.*

List of speakers who are nonmembers: None

Chapter 22

ASA Program Chair Discretion (includes Memorial Sessions) (4)

Program Chair: **Scott, David W.**

22.1 asa.prog.chair.01

Sponsoring Section/Society: ASA Program Chair

Session Slot: 4:00- 5:50 Sunday
Estimated Audience Size: 100-150
AudioVisual Request: Two Overheads

Session Title: Donald W. Marquardt Memorial Session

Theme Session: No
Applied Session: No

Session Organizer: **Skwish, Joseph N.** DuPont Company

Address: 5123 New Kent Road, Wilmington DE 19808
Phone: (302)695-1784
Fax: (302)695-2747

Email: pqmjns@pccvax.dnet.dupont.com

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 5 minutes First Speaker - 25 minutes Second Speaker - 25 minutes Third Speaker - 25 minutes Fourth Speaker - 25 minutes Floor Discusion - 10 minutes

Session Chair: **Skwish, Joseph N.** DuPont Company

Address: 5123 New Kent Road, Wilmington, DE 19808
Phone: (302)695-1784
Fax: (302)695-2747
Email: pqmjns@pccvax.dnet.dupont.com

1. Nonlinear Estimation

Bates, Douglas M., University of Wisconsin

Address: Statistics, University of Wisconsin 1210 W. Dayton St. Madison, WI 53706-1685
Phone: (608)262-2598
Fax: (608)262-0032
Email: bates@stat.wisc.edu

Abstract: One of the turning points in the history of nonlinear estimation was the publication of Donald Marquardt's 1963 SIAM paper "An Algorithm for Least Squares Estimation of Nonlinear Parameters". This paper has achieved the status of a citation classic, being listed as the third most frequently cited paper in all the mathematical sciences. It was a product of the best kind of interplay of statistical theory and practice. The ideas in this paper grew from Don Marquardt's day-to-day consulting experience, his search for the causes of failure in the nonlinear least squares routines being used at the time, and his clever insight into how to repair these deficiencies. It is not an accident that this paper is a classic. It has spawned whole areas of research in nonlinear estimation and in numerical analysis. We review the basis for the methodology Don Marquardt introduced and give some explanation of his method and its ramifications.

2. Contributions to ASA

Bailar, Barbara A., National Opinion Research Center

Address: VP for Survey Research, National Opinion Research Center, 1155 E 60th St., Chicago, IL 60637
Phone: (773)256-6070
Fax: (773)753-7886
Email: bailar-b@norcmail.uchicago.edu

Abstract: *Don Marquardt played a major role in ASA that went far beyond his presidency. As the major designer of strategic planning for ASA in the eighties, Don set the stage for making changes in the ASA to bring it more in line with member needs. He was also an important player in decisions about ASA owning a building for its operations. However, Don's most important contribution to ASA was the example he set as a consulting statistician who always took the personal equation into account.*

3. Contribution to International Quality Standards

Peach, Robert W., Robert Peach & Associates

Address: 200 West Cornwall Rd. #126, Cary NC 27511-3802
Phone: (919)319-1982
Fax: (919)319-1984
Email: robertpeach@compuserve.com

Abstract: *Don's background at DuPont made him the perfect choice to take part in the process of crafting a family of international quality standards. From its beginning in 1980, he was a member of the Technical Committee that saw the opportunity and need for harmonized quality system standards. Don chaired the writing of ISO 9000, the guidance document describing the application of the ISO 9000 family of standards. Upon the passing of Richard Freund, Don assumed the role of leader of the US delegation. Don was asked to chart the direction that ISO 9000 development should take, leading to the publication, in 1991, of*

the document "vision 2000," prepared by an international group that Don chaired. Even today, the major actions that are now in motion to adapt and improve the ISO 9000 family of quality system standards continue to reflect Don's guidance and vision.

4. Overview of Don Marquardt's DuPont Career

Pfeifer, Charles G., DuPont Company

Address: DuPont Engineering, DuPont Bldg 2086-1, 11th & Market St., Wilmington, DE 19898
Phone: (302)774-2404
Fax: (302)774-2308
Email: pfeifecg@engg.dnet.dupont.com

Abstract: Don Marquardt's career with Dupont spanned 39 years. During that time he was a major player in advancing the use of statistics and increasing the influence that it had in DuPont R&D, Manufacturing, Marketing and other areas of application. His contributions came about not only through his own direct involvement in applications, but also through mentoring of other statisticians and scientists alike. He guided the development of internal statistical software and statistics training courses, and built and managed one of the largest groups of statisticians in the process industries. Much of the technical work on which he collaborated required advances which he developed and published externally. He was tireless in devoting his own personal time to increasing his knowledge of such fields as strategic planning and manufacturing quality improvement standards. He leveraged this knowledge for further synergistic value and benefit to DuPont and to the broader society with which he was involved.

List of speakers who are nonmembers: Robert W. Peach

22.2 asa.prog.chair.02

Sponsoring Section/Society: ASA Program Chair Session

Session Slot: 10:30-12:20 Thursday
Estimated Audience Size: xx-xx
AudioVisual Request: TBA

Session Title: Memorial Session: James T. Massey

Theme Session: No
Applied Session: Yes

Session Organizer: **Zell, Elizabeth R.**

Session Organizer: **Ezzati-Rice, Trena M.** CDC

Address: 1600 Clifton Road, MS-E52
Phone: 404-639-8392
Fax:
Email: ezr1@cdc.gov and tme1@cdc.gov

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 5 or 0 minutes First Speaker - 30 minutes (or 25) Second Speaker - 30 minutes Third Speaker - 30 minutes Discussant - 10 minutes (or none) Floor Discusion - 10 minutes (or 5 or 15)

Session Chair: **Zell, Elizabeth R.** CDC

Address: 1600 Clifton Road, MS-E62
Phone: 404-639-8392
Fax:
Email: ezr1@cdc.gov

1. Using the National Health Interview Survey as a Sampling Frame for Other Health-Related Surveys

Ezzati-Rice, Trena M., NCHS

Address: 6525 Belcrest Road; Hyattsville, Maryland 20782
Phone: 301-436-7022 x133
Fax: 301-436-7955

Email: tme1@cdc.gov
Cohen, Steven, AHCPR

Abstract: A number of important health and health care surveys are carried out by the Department of Health and Human Services (DHHS). A major priority of the DHHS Data Council over the past several years has been the movement toward integrated, coordinated data systems that can meet multiple health data needs in an efficient and linked framework. The linkage of data systems through the use of common core questionnaires and sampling frames is designed to enhance analytic capabilities, maximize survey efficiencies, and minimize survey costs. The National Health Interview Survey (NHIS), a continuous survey of the U.S. population conducted by the National Center for Health Statistics, is an important component of the nation's health statistics data collection system. DHHS conducts a number of other health and health-related surveys. This paper will highlight the key features of linking the design of three major population-based surveys to that of the NHIS. The use of NHIS samples from prior years as screening samples to find domains of special interest is one advantage of linkage. Cycle IV of the National Survey of Family Growth (NSFG) was the first periodic household survey to use the NHIS as a sampling frame. The NSFG, a survey of women ages 15 to 44 with oversampling of Hispanic and black women, is designed to produce estimates of factors related to pregnancy and the health of women and infants. A redesigned National Health and Nutrition Examination Survey (NHANES IV) will move to a continuous survey and will use NHIS as a sampling frame. Finally, an important longitudinal panel survey on insurance, expenditures, health care utilization, and sources of payment, The Medical Expenditure Panel Survey (MEPS), conducted by the Agency for Health Care Policy Research, also uses the NHIS as its sampling frame. The advantages, as well as some of the issues and limitations associated with the design of these linked surveys, such as respondent burden, tracking, response rates and use of the NHIS sample for multiple surveys will also be discussed. Finally, we will outline potential methodological evaluations for the next redesign of the

NHIS as it relates to linked surveys in the future.

2. Telephone Surveys: Past, Present, and Future

Cox, Brenda G., Mathematica Policy Research, Inc.

Address: 600 Maryland Ave SW, Suite 550; Washington, D.C. 20024
Phone: 202-484-4231
Fax: 202-863-1763
Email: BCox@Mathematica-Mpr.com
Mitchell, Susan, Mathematica Policy Research, Inc.

Abstract: The past 20 years has seen the telephone become the predominate data collection mode for publically and privately sponsored surveys. Paying special attention to the contributions made by James Massey, this paper describes critical milestones in the recognition of telephone surveys as a valid data collection approach and the challenges becoming apparent for the future. Massey and Thornberry were one of the first to document the dramatic improvement in telephone coverage for the U. S. population. This improved telephone coverage together with its low cost led survey sponsors to substitute the telephone for more expensive personal interviews or less reliable mail questionnaires. Simultaneously, researchers initiated investigations designed to identify ways to improve the quality of telephone survey data. Methodological research was conducted to determine methods to improve response rates, to describe mode differences between telephone and face-to-face surveys, and to delineate response category ordering effects. This research together with the widespread adoption of computer-assisted telephone interviewing (CATI) in the late 1970's and early 1980's led to continuing improvement in the quality of telephone survey data. The last decade has seen increasing emphasis on reducing the cost of telephone interviews while maintaining data quality. Methodologists now focus on issues such as optimal calling times and patterns, the effect of interviewer and respondent incentives, and automated approaches for call scheduling and dialing. Samplers have shifted from the classic Mitofsky-Waksberg approach for random digit dialed surveys to list-assisted sampling, reducing survey costs while making

data collection easier to manage. As our final topic, we describe what we see as future directions for telephone surveys. Samplers need to explore the use of telephone surveys as an efficient screening tool for rare populations and, in combination with face-to-face interviews, as an affordable method when state-level estimation of low income populations is needed. Methodologists need to pursue parallel investigations to uncover methods to overcome interviewing barriers posed by innovations such as answering machines and automatic caller identification as well as how to take advantage of new technologies such as touchpad interviewing, predictive dialing, and decentralized interviewing.

3. An Evaluation of the Bias in Telephone and Face-to-Face Interviewing for the National Health Interview Survey

Biemer, Paul P., Research Triangle Institute

Address: 3040 Cornwallis Road; Post Office Box 12194; Research Triangle Park, North Carolina 27709
Phone: 919-541-6602
Fax: 919-541-7431
Email: ppb@rti.org

Abstract: This study examines the quality of RDD data relative to data collected by face to face interviewing in the National Health Interview Survey (NHIS). The motivation for the study is to determine the feasibility of combining RDD and NHIS data in order to reduce the variance of state-level estimates. Currently, since NHIS is strictly a national survey, there are only a few states where the annual NHIS sample sizes are adequate to produce state-level estimates of the required precision. If this supplementation of the NHIS sample with state-based RDD samples proves to be feasible, dual frame survey estimates of NHIS characteristics could be an option for states who need state-level estimates of greater precision than can be produced by the NHIS.

The primary determinant of the feasibility of this dual frame, dual mode approach for the NHIS is the relative quality of the RDD data. The present study was designed to evaluate the relative magnitudes of the nonsampling error components associated

with the RDD survey and the NHIS. Two states were selected for the study: Texas and California. In each of these states, about 2000 interviews were conducted by CATI and an equal number of interviews were conducted for the regular NHIS. Further, about 500 RDD households and 1000 NHIS households were reinterviewed in each state within three weeks of the original interview using replicate reinterview procedures. Finally, a follow-up of NHIS nonrespondents was conducted netting a total of 214 respondents (40% nonresponse conversion rate).

In this paper, we describe the results of a latent class analysis of the combined NHIS and RDD/NHIS data. Estimates of measurement and nonresponse bias are presented and compared for both modes. Using these estimates, we compare the biases of RDD CATI and face to face administration for the NHIS. We conclude that, when the effects of nonresponse bias and measurement bias are combined, the data quality for the telephone mode does not differ appreciably from that of face to face interviewing for key NHIS data items. Our general results have implications for most surveys conducted by centralized, computer assisted telephone interviewing.

Discussant: **White, Andrew A.** National Research Council

Address: 2101 Constitution Ave. N.W., Rm. HA192 Washington, D.C. 20418
Phone: 202-334-3096
Fax: 202-334-3751
Email: awhite@nas.edu
List of speakers who are nonmembers: None

22.3 asa.prog.chair.03

Sponsoring Section/Society: ASA, IMS, IISA

Session Slot: 2:00- 3:50 Monday
Estimated Audience Size: xx-xxx
AudioVisual Request: None

Session Title: *R. R. Bahadur Memorial Session*

Theme Session: No
Applied Session: No

Session Organizer: **Stigler, Stephen M.** University of Chicago

Address: Department of Statistics 5734 University Ave. Chicago, IL 60637
Phone: 773-702-8328
Fax: 773-955-5893
Email: stigler@galton.uchicago.edu

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 5 minutes First Speaker - 30 minutes Second Speaker - 30 minutes Third Speaker - 30 minutes Floor Discusion - 10 minutes

Session Chair: **Stigler, Stephen M.** University of Chicago

Address: Department of Statistics 5734 University Ave. Chicago, IL 60637
Phone: 773-702-8328
Fax: 773-955-5893
Email: stigler@galton.uchicago.edu

1. Bahadur and Sufficiency

Diaconis, Persi, Cornell University

Address: Mathematics Department White Hall Cornell University Ithaca, NY 14853
Phone:
Fax:
Email: ims@math.cornell.edu

Abstract: *Bahadur's work on sufficiency illuminated sequential analysis, Bayesian decision theory, and mathematical statistics generally.*

2. Consistency Issues For Maximum Likelihood Estimates and Bayesian Analysis

Ghosh, Jayanta K., ISI and Purdue University

Address: Department of Statistics Purdue University West Lafayette, IN 47907, and Theoretical Statistics and Mathematics, ISI, 203 B.T. Road, Calcutta 700 035, INDIA

Phone:

Fax:

Email: ghosh@stat.purdue.edu

Ramamoorthi, R.V., ISI

Ghosal, S., ISI

Abstract: Bahadur was the first to raise in 1958 the question of consistency of the maximum likelihood estimate of the true density and later provided a definitive answer through a theorem and counterexamples. His treatment was nonparametric in spirit. Some work has been done recently on the convergence of non parametric maximum likelihood estimates, based on empirical processes. There are related questions of consistency of the posterior in non parametric infinite dimensional Bayesian analysis, as in Diaconis and Freedman (Annals of Statistics, 1986). In this area too, for example in Bayesian density estimation, there have been significant new developments. My review will begin with a look at Bahadur's work and then provide a relatively non-technical and self contained survey of recent work.

3. Bahadur and Asymptotic Efficiency Theory

Wong, Wing Hung, UCLA

Address: UCLA Interdivisional Program in Statistics 8118 Mathematical Sciences, 155404 Box 951554 Los Angeles, CA 90095-1554

Phone:

Fax: (310) 206-5658

Email: whwong@stat.ucla.edu

Shen, Xiaotong, Ohio State University

Abstract: Raj Bahadur had made fundamental contributions to the study of asymptotic efficiency in estimation and testing. We review two aspects of his impact in this area. First, his paper in 1964 clarified the best possible performance for estimators that possess asymptotic distributions. This work represented an elegant completion of a series of deep studies on this issue by Edgeworth, Fisher and LeCam. A second and perhaps more profound contribution by Bahadur was an alternative approach to asymptotics based on large deviation probability. His 1960 paper introduced global and local bounds for the best possible rate, known as Bahadur efficiency, by an application of the Neyman-Pearson Lemma. This work had stimulated much subsequent research by many other investigators. It was a great leap that added immensely to our understanding of many fundamental issues in estimation and testing.

List of speakers who are nonmembers: None

22.4 asa.prog.chair.04

Sponsoring Section/Society: ASA Program Chair

Session Slot: 10:30-12:20 Monday
Estimated Audience Size: 250
AudioVisual Request: None

Panel Session Title: The Role of Statisticians in Public Policy

A panel discussion of issues related to this year's Joint Statistical Meeting theme, as proposed by the ASA President, Professor David Moore. (He'll be working on this more this week.)

Theme Session: Yes
Applied Session: No

Session Organizer: **Moore, David S.** Purdue University

Address: Dept Of Statistics Purdue University W Lafayette, IN 47907-1399
Phone: 317-494-6050
Fax: 317-494-0558
Email: dsm@stat.purdue.edu

Session Organizer: **Scott, David W.** Rice University

Address: Dept of Statistics, MS-138 Rice University 6100 Main St Houston, TX 77005-1892
Phone: 713-527-6037
Fax: 713-285-5476
Email: scottdw@rice.edu

Session Timing: 110 minutes total (Sorry about format):

> Opening Remarks by Chair - 5 or 0 minutes First Speaker - 30 minutes (or 25) Second Speaker - 30 minutes Third Speaker - 30 minutes Discussant - 10 minutes (or none) Floor Discussion - 10 minutes (or 5 or 15)

Session Chair: **Moore, David S.** Purdue University

Address: Dept Of Statistics Purdue University W Lafayette, IN 47907-1399
Phone: 317-494-6050
Fax: 317-494-0558
Email: dsm@stat.purdue.edu
Stark, Philip B. University of California, Berkeley

Address: Deptartment of Statistics University of California Berkeley, CA 94720-3860
Phone: 510-642-1430
Fax: 510-642-7892
Email: stark@stat.berkeley.edu
Fienberg, Stephen E. Carnegie Mellon University

Address: Carnegie Mellon University, Dept. of Statistics, Pittsburgh, PA 15213

Phone: 412-268-2723
Fax: 412-268-7828
Email: fienberg@stat.cmu.edu
Thompson, James, R. Rice University

Address: Dept of Statistics, MS-138, Rice University, P.O. Box 1892, Houston, TX 77005-1892
Phone: 713-527-4828
Fax: 713-285-5476
Email: thomp@rice.edu
Wallman, Katherine Office of Management and Budget

Address: Office of Management and Budget, Washington, D.C. 20603
Phone: 202-395-3093
Fax:
Email: WALLMAN_K@a1.eop.gov
List of speakers who are nonmembers: None

Chapter 23

Biometric Society (ENAR & WNAR) (12)

Program Chair: **Davidian, Marie** Program Chair: **Stram, Daniel**

23.1 biometric.soc.01

Sponsoring Section/Society: ENAR

Session Slot: 8:30-10:20 Tuesday
Estimated Audience Size: 125-175
AudioVisual Request: Two Overheads

Session Title: Advances and Applications of Model Averaging and Mixtures

The use of mixtures (or normals, for example) and model averaging forms an important class of tools for analysis of complex data sets and accounting for model uncertainty. This session will describe novel applications of and recent advances in mixture modeling and Bayesian model averaging in biometrics.

Theme Session: No
Applied Session: Yes

Session Organizer: **Carroll, Raymond J.** Texas A&M University

Address: Department of Statistics Texas A&M University College Station TX 77843-3143
Phone: 409-845-3141
Fax: (409-845-3144
Email: carroll@stat.tamu.edu

Session Timing: 110 minutes total (Sorry about format):

> Opening Remarks by Chair - 0 minutes First Speaker - 25 minutes (or 25) Second Speaker - 25 minutes Third Speaker - 25 minutes Fourth Speaker - 25 minutes Floor Discusion - 10 minutes

Session Chair: **Berry, Scott** Texas A&M University

Address: Department of Statistics Texas A&M University College Station TX 77843-3143
Phone: 409-845-3141
Fax: (409-845-3144
Email: berry@stat.tamu.edu

1. Bayesian Model Averaging and Other Stategies in Case-Control Studies

Raftery, Adrian E., University of Washington

Address: Dept of Statistics University of Washington B313 Padelford Hall Box 354322 Seattle WA 98195-4322
Phone: 206-543-4505
Fax: 206-685-7419
Email: raftery@stat.washington.edu
Viallefont, Valerie, INSERM, France
Richardson, Sylvia, INSERM, France

Abstract: Hundreds of observational case-control studies are reported each year in the epidemiological and medical literatures. A key issue in the analysis of such studies is the choice of control variables. Typically, many possible control variables are considered (on the order of several dozen). The most used strategies have been stepwise logistic regression, and the two-stage method *of Mickey and Greenland (1989). With these strategies, inference about treatment effects (estimates, confidence intervals, tests and P values) is made conditionally on the selected set of control variables. We argue that this may be misleading because it ignores model uncertainty, which can be considerable.*

We propose instead the use of Bayesian model averaging (BMA), using a prior distribution for the parameters that is based on a sample of recent studies from the American Journal of Epidemiology. The software GLIB (http://lib.stat.cmu.edu/S/glib) is used (Raftery, 1996, Biometrika). We report on a simulation study, designed to mimic the features of commonly encountered studies in the literature. We find that conditional P values from both the stepwise and two-stage methods greatly overstate the evidence for an effect, while BMA posterior probabilities reflect it reasonably accurately. Also, the BMA point estimates have lower mean squared errors than those ¿from either stepwise or two-stage variable selection. We conclude that any method that selects a single model can be misleading, and that it is important to take account of model uncertainty when making inference from case-control studies.

2. Does Particulate Matter Particularly Matter

Clyde, Merlise, Duke University

Address: Institute of Statistics and Decision Science Duke University Box 90251 Durham NC 27708-0251

Phone: 919-681-8440

Fax: 919-681-8594

Email: clyde@isds.duke.edu

Abstract: The effect of particulate matter on mortality is currently an important issue, as the U.S. Environmental Protection Agency has proposed new standards. Many of the epidemiological models are based on Poisson regression models for the mortality counts, with independent variables including various daily and lagged meteorological variables and measures of particulate matter. The high correlations among many of the explanatory *variables makes tradition model selection difficult. The statistical significance of PM10 (particulate matter of aerodynamic diameter less than 10 micrometers), however, depends on the choice of meteorological variables that are selected and the choice of PM10 variables. Issues of variable choice and model uncertainty due to variable selection thus can play an important role in making decisions. We use Bayesian model averaging to assess the effect of PM10 on mortality, taking into account model uncertainty about which meteorological and PM10 variables should be included. Because there are a large number of possible models with over two hundred variables, we introduce an approximation to the posterior distribution that allows for efficient Markov Chain Monte Carlo sampling of models with high posterior probability from high dimensional model spaces. We present posterior distributions under model averaging of the effect of PM10 on mortality and of the relative risk and contrast these results with previous analyses. The methods are applicable to a wide range of generalized linear models where prediction and model uncertainty are important issues.*

3. Flexible Bayesian and Frequentist Inference Using Mixtures

Wasserman, Larry, Carnegie Mellon University

Address: Department of Statistics Baker Hall, 229I Carnegie Mellon University Pittsburgh PA 15213
Phone: 412-268-8727
Fax: 412-268-7828
Email: larry@stat.cmu.edu
Carroll, Raymond J., Texas A&M University
Roeder, Kathryn, Carnegie Mellon University

Abstract: Mixtures of Normals models provide a simple, yet flexible family of distributions. First we'll discuss these general issues: problems with improper priors, selecting the number of components, and computational considerations. Then we'll describe a few applications including: density estimation, measurement models and combining studies. Time permitting, we'll also discuss a comparison of sieves (where the number of components

grows with sample size) versus infinite mixtures (such as Dirichlet process mixtures).

4. Analyzing the Effects of Haplotype Variation on a Continuous Variable Using Bayesian Model Averaging

Iturria, Stephen, Texas A&M University

Address: Department of Statistics Texas A&M University College Station TX 77843-3143
Phone: 409-845-3141
Fax: (409-845-3144
Email: iturria@stat.tamu.edu

Abstract: Conventional approaches to model fitting make inferences on model characteristics under the assumption that the chosen model is correct. There can be much uncertainty, however, in the correctness of the selected model. The idea behind Bayesian model averaging (BMA) is that when possible, one should account for this uncertainty when making inferences. When inference on a quantity, Z, common to several competing models is desired, BMA attempts to account for model uncertainty by characterizing our knowledge about Z with a composite of posterior distributions for Z, each computed under a different model. The composite is a weighted average, with weights taken as the posterior probabilities for some subset of the models being considered. In this presentation we describe how one could use BMA to analyze the effect of variation in haplotypes of the X gene on observed levels in Y.

List of speakers who are nonmembers: None

23.2 biometric.soc.02

Sponsoring Section/Society: ENAR

Session Slot: 2:00- 3:50 Monday
Estimated Audience Size: 300

AudioVisual Request: Two Overheads

Session Title: Statistical Models for Longitudinal Studies: Unifying Longitudinal and Survival Analysis Methods

Description of session: Data collected in studies in public health routinely include both observations of times to events of interest and other measures taken repeatedly over time; e.g. in AIDS clinical trials, both time to death or other event and longitudinal measures of CD4 count, viral load, and other immunological or virological variables are observed. This session focuses on recent advances in methods of analysis that exploit information of both types.

Theme Session: No
Applied Session: Yes

Session Organizer: **Hogan, Joseph** Brown University

Address: Center for Statistical Sciences Box G-H Brown University Providence RI 02912
Phone: 401-863-9243
Fax: 401-863-9182
Email: jhogan@stat.brown.edu

Session Timing: 110 minutes total (Sorry about format):

> Opening Remarks by Chair - 5 minutes First Speaker - 25 minutes Second Speaker - 25 minutes Third Speaker - 25 minutes Discussant - 20 minutes Floor Discusion - 10 minutes

Session Chair: **Hogan, Joseph** Brown University

Address: Center for Statistical Sciences Box G-H Brown University Providence RI 02912
Phone: 401-863-9243
Fax: 401-863-9182
Email: jhogan@stat.brown.edu

1. The Analysis of Longitudinal Data – An Expanded Perspective

Zeger, Scott, Johns Hopkins University

Address: Department of Biostatistics Johns Hopkins University 615 N. Wolfe Street, Room E3132 Baltimore MD 21205-2179
Phone: 410-955-3067
Fax: 410-955-0958
Email: szeger@jhsph.edu

Abstract: *Longitudinal studies in public health permit observations of both times to key health events as well as repeated measures of health outcomes. Statistical methods for times-to-events (survival analysis) and for repeated events (longitudinal data analysis) have developed along independent lines even though substantive questions can often be addressed through joint analysses of both kinds of outcomes.*

In this talk, we consider opportunities for unifying statistical methods for times-to-events and repeated measures under an expanded definition of longitudinal data analysis. The ideas will be illustrated with examples from public health research.

2. Nonparametric Incomplete Data Estimation of Stochastic Process Structures Related to Survival Models

Jewell, Nick, University of California, Berkeley

Address: Department of Statistics University of California at Berkeley 367 Evans Hall #3860 Berkeley CA 94720-3860
Phone: 510-642-5361
Fax: 510-642-7892
Email: jewell@stat.Berkeley.edu

Abstract: *Considerable attention has been given to nonparametric estimation of simple survival models based on data subject to random censoring and truncation. Recently, there has been additional interest in similar problems based on current status data. In this talk, we will discuss extensions of some of these ideas by considering survival data in a stochastic process setting.*

The extensions we have in mind are to (i) multidimensional survival processes, and (ii) more general stochastic processes than the familiar two-state survival models. The first of these extensions includes consideration of so-called marker processes where one component of the process is a simple survival counting process and other components represent marker variables thought to be associated with survival.

3. Non-ignorable Non-response in Longitudinal Studies with Continuous Drop-out Times

Scharfstein, Dan, Johns Hopkins University

Address: Department of Biostatistics Johns Hopkins University 615 N. Wolfe Street, Room E3132 Baltimore MD 21205-2179
Phone: 410-955-3067
Fax: 410-955-0958
Email: dscharf@athena.biostat.jhsph.edu
Robins, James M., Harvard School of Public Health

Abstract: Consider a longitudinal study in which we follow subjects for T months after enrollment. At this point, we measure the primary endpoint of interest, Y. In addition to this measurement, a series of baseline covariates, $V(0)$, and time dependent covariates, $V(t)$, are collected. We are interested in making inference about the mean of Y, μ. Unfortunately, some individuals drop-out before time T. If we let Q denote the continuous drop-out time, then the observable data for an individual is $(X, \Delta, \bar{V}(X), \Delta Y)$, where $X = \min(Q,T)$, $\Delta = I(Q \geq T)$, and $\bar{V}(t) = \{V(u) : 0 \leq u \leq t\}$ is the covariate history through time t. If Q is unrelated to Y, then Y is said to be missing completely at random and the sample average of the observed outcomes is an unbiased estimate of μ.

The purpose of this paper is to show how to make inferences about μ when the drop-out time is modeled semiparametrically, no restrictions are placed on the joint distribution of the outcome and covariates, and the outcome is not missing completely at random.

In particular, we assume that the law of Q given $\bar{V}(T)$ and Y follows a proportional hazards model of the form

$$\lambda_{Q|\bar{V}(T),Y}(u|\bar{V}(T), Y) = \lambda(u)\exp(\delta' h(\bar{V}(u)) + \alpha^* Y)$$

where $\lambda_{Q|\bar{V}(T),Y}(u|\bar{V}(T), Y)$ is the hazard of Q given $\bar{V}(T)$ and Y, $\lambda(u)$ is some arbitrary, non-negative function of u, $h(\cdot)$ is a r-dimensional function of the covariate history, δ is an rx1 vector of unknown parameters, and α^* is a known constant. Note that $\alpha^* = 0$ corresponds to missing at random and $\alpha^* \neq 0$ corresponds to non-ignorable non-response. We have chosen to fix this latter constant because we know that it is at best difficult to estimate. Instead, we perform a sensitivity analysis to see how inference about μ changes as we vary α^* over a plausible range of values. We apply our approach to the analysis of ACTG 175, an AIDS clinical trial in which we define the primary outcome as CD4 at 56 weeks after enrollment. Repeated measurements on CD4 counts act as the time dependent covariates, while gender, age, race, and information on other risk factors serve as the baseline covariates.

Discussant: **Wei, L.J.** Harvard School of Public Health

Address: Department of Biostatistics Harvard School of Public Health 677 Huntington Avenue Boston MA 02115
Phone: 617-432-2826
Fax: 617-739-1781
Email: wei@hsph.harvard.edu
List of speakers who are nonmembers: None

23.3 biometric.soc.03

Sponsoring Section/Society: ENAR

Session Slot: 10:30-12:20 Tuesday
Estimated Audience Size: 125-175
AudioVisual Request: Two Overheads

Session Title: Statistical Methods for Nonignorable Missing Data

Description of the session: Policy makers rely on accurate analysis of data to make appropriate decisions. However, when we report statistical results to policy makers, we are seldom telling them what the assumptions behind the results are and what the effects of violating these assumptions on the conclusions are. One example is statistical analysis of missing data. Most commonly used methods for analyzing missing-data require the missing-at-random (or ignorable) assumption. However, if the missing-data mechanism is non-ignorable, what are the effects of assuming ignorable missing-data mechanism on the conclusions and what are some alternative approaches when the missing-data mechanism is non-ignorable? In this session, we will have three talks. The first talk, by Dr. Rotnitzky, will focus on semiparametric sensitivity analysis with non-ignorable non-response data; the second talk, by Dr. Baker, will focus on The design and analysis of a randomized trial with a missing binary outcome: the role of an auxiliary variable; the third talk, by Dr. Zhou, will focus on statistical methods for evaluation of diagnostic accuracy when verification is subject to non-ignorable missing-data mechanism. The discussant in the session will be Dr. Daniel Heitjan at Columbia University.

Theme Session: No
Applied Session: No

Session Organizer: **Zhou, Xiao-Hua (Andrew)** Indiana University School of Medicine

Address: Division of Biostatistics Indiana University School of Medicine Riley Research Wing, 135 699 West Drive Indianapolis IN 46202-5119
Phone: 317-274-2696
Fax: 317-274-2678
Email: zhou@mako.biostat.iupui.edu

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 5 minutes First Speaker - 30 minutes Second Speaker - 30 minutes Third Speaker - 30 minutes Discussant - 10 minutes Floor Discussion - 5 minutes

Session Chair: **Rubin, Donald B.** Harvard University

Address: Department of Statistics Harvard University 1 Oxford Street Cambridge MA 02138-2901
Phone: 617-495-5498
Fax: 617-496-8057
Email: rubin@stat.harvard.edu

1. A Note About Multiple Imputation

Wang, Naisyin, Texas A&M University

Address: Department of Statistics Texas A&M University College Station TX 77843-3143
Phone: 409-845-3141
Fax: (409-845-3144
Email: nwang@stat.tamu.edu
Robins, Jamie, Harvard University

Abstract: We consider the asymptotic behaviour of various parametric multiple imputation procedures. The asymptotic variance structure of the resulting estimators is provided. This result is used to compare the relative efficiencies of different imputation procedures and it applies to non-ignorabl missing mechanism provided that certain conditions are fullfilled. Some examples are used to illustrate the results.

2. The Design and Analysis of a Cancer Prevention Trial: Adjusting for a Missing Binary Outcome, an Auxiliary Variable, and All-or-None Compliance

Baker, Stuart G., National Cancer Institute

Address: Biometry Branch Division of Cancer Prevention National Cancer Institute EPN 344 9000 Rockville Pike Bethesda MD 20892
Phone: 301-496-7708
Fax: 301-402-0816
Email: sb16i@nih.gov

Abstract: *The National Cancer Institute is recruiting subjects into a randomized trial of daily finasteride versus placebo. The outcome is prostate cancer incidence determined by biopsy, either after seven years or following an elevated PSA (prostate-specific antigen) on yearly screening. There are two complications in the analysis. First, missing a biopsy will likely depend on whether or not a subject has an elevated PSA. The indicator of elevated PSA is called an auxiliary variable because it is observed after randomization and prior to endpoint. Second, there will likely be contamination in the placebo arm and nonadherence in the finasteride arm, starting soon after randomization. Compliance in this situation is called all-or-none. To adjust for these complications we formulate the appropriate likelihoods and obtain closed-form maximum likelihood estimates and variances. Some of these likelihoods involve nonignorable missing-data mechanisms.*

3. Statistical Methods for Evaluation of Diagnostic Accuracy When Verification is Subject to Non-Ignorable Missing Data Mechanism

Zhou, Xiao-Hua (Andrew), Indiana University School of Medicine

Address: Division of Biostatistics Indiana University School of Medicine Riley Research Wing, 135 699 West Drive Indianapolis IN 46202-5119
Phone: 317-274-2696
Fax: 317-274-2678
Email: zhou@mako.biostat.iupui.edu

Abstract: *The accuracy of a diagnostic test is often measured by its sensitivity, specificity, positive predictive and negative predictive values. More generally, a receiver operating characteristic (ROC) curve may be used to represent the accuracy of a diagnostic test. To calculate these measures, we need to determine the disease status of each patient in the sample. The procedure that establishes the patient's disease status is referred to as a gold standard. However, for some studies, only a subset of the patients with diagnostic test results are chosen to receive the gold standard assessment. For example, in some clinical circumstances,*

the disease status is obtained by surgery or biopsy, usually when a diagnostic test has provided strong evidence of the presence of disease. If the study population consists of only verified cases, the estimated accuracy of the diagnostic test may be biased. This type of bias is called verification bias, and may be treated as a missing-data problem. In this talk, I will first review available verification bias correction methods under the ignorable assumption. Then, I will discuss the the effects of non-ignorable verification bias on estimates, derived under the ignorable assumptions, and discuss a test procedure for the ignorable assumption. Finally, I will illustrate the methods with datasets from several clinical studies.

Discussant: **Heitjan, Daniel** Columbia University

Address: Division of Biostatistics Columbia University School of Public Health New York NY 10032-3702
Phone: 212-305-9557
Fax: 212-305-9408
Email: dfh5@columbia.edu
List of speakers who are nonmembers: None

23.4 biometric.soc.04

Sponsoring Section/Society: ENAR

Session Slot: 10:30-12:20 Wednesday
Estimated Audience Size: 60-80
AudioVisual Request: Two Overheads

Session Title: Affecting Policy through Environmental Statistics

Session Organizer: **Ryan, Louise M.** Harvard University

Address: DFCI, 44 Binney Street, Boston Ma 02115
Phone: 617-632-3602
Fax: 617-632-2444
Email: ryan@jimmy.harvard.edu

Session Timing: 110 minutes total (Sorry about format):

Chair - 5 minutes First Speaker - 25 minutes (including 3 minutes questions) Second Speaker - 25 minutes Third Speaker - 25 minutes Discussant - 15 minutes Floor Discusion - 15 minutes

Session Chair: **Young, Linda J.** University of Nebraska

Address: Department of Biometry U of Nebraska Lincoln NE 68583-0712
Phone: 402 472 2903
Fax: 402 483 2392
Email: biom025@unlvm.unl.edu

1. Dose-Rate Considerations in Risk Assessment: Statistical Models and Policy Implications

Williams, Paige L., Dept. of Biostatistics, Harvard School of Public Health

Address: Biostatistics, 677 Huntington Avenue Boston Ma 02115
Phone: (617) 432-3872
Fax: (617) 432-2832
Email: paige@hsph.harvard.edu

Abstract: Regulatory agencies have become increasingly aware of the need to develop standards for a wider range of exposure conditions relevant for toxicity endpoints, particularly varying lengths of exposure. The strategy for comparing responses among exposures of different durations has often relied on a conjecture proposed by Haber (1924) that response levels should be equivalent for any constant multiple of dose times duration, i.e., whenever the cumulative exposure remains constant. More flexible statistical models have been developed for assessing joint effects of dose and duration, yet these typically require larger and more complex study designs. Current default approaches for setting regulatory standards in the context of dose-rate exposures will be reviewed, and implications of more flexible approaches on regulatory policy will be addressed.

2. Mechanistic Modeling in Environmental Policy: Challenges for Statistics

Portier, Christopher J., NIEHS

Address: NIEHS PO Box 12233 MD A3-06 Research Triangle Park, NC 27709
Phone: 919 541 4999
Fax: 919-541-1479
Email: portier@niehs.nih.gov

Abstract: There is now underway a concerted effort by the regulatory community to "improve" decisions about environmental health risks through the use of "mechanistic models" explaining the relationship between exposure and risk. The benefit from the use of these models is an increased linkage of a broad array of data sources and the u se of historical understanding of the etiology of a toxicity in the formulation of the model. However, these efforts result in a difficult task for statisticians i n terms of definition (what constitues "mechanistic modeling" as compared to an al ternative), parameter estimation (identifiability, likelihood construction, varian ce estimation, etc) and in validation of a highly complicated modeling scheme. Th is talk addresses some of these concerns and suggests avenues of research for stat isticians to improve the use of "mechanistic models" in risk assessment.

3. Regulating the Effects of Commercial Salmon Net Fisheries on Protected Seabirds

Conquest, Loveday L., Associate Dean And Professor, College Of Ocean & Fishery Sciences Univ. Of Washington

Address: Box 3355645 Seattle WA 98195-5645
Phone: 206 543 1708
Fax: 206 543 6263
Email: conquest@cqs.washington.edu
Melvin, Edward F., University of Washington, Washington Sea Grant Program

Abstract: In developing regulations for fishing gear and fishing techniques that reduce the accidental entanglement of endangered seabirds, we compared entanglement rates and fish catch rates in different types of salmon fishnets, and across different times of day. We used loglinear models to analyze both bird entanglement and fish catch data. Results indicate that seabird entanglement and salmon catch rates vary significantly among fishnet types and times of day. Particular patterns of variation were specific to the type of bird being investigated. Common Murres and Rhinoceros Auklets responded differently (in terms of entanglement rates) to different types of nets, and depending upon the time of day. This can be problematic when trying to devise fishing regulations across a broad array of circumstances. Fishing policy changes based on our work involve use of visual net barriers, acoustic pingers for fishnets, and daytime only fishing to reduce endangered seabird bycatch by commercial salmon fishing boats.

Discussant: **Ross, N. Phillip** Director for the Center for Environental Statistics, USEPA

Address: 401 M Street, SW Room 3101 Mail Stop 2163 Washington DC 20450
Phone: 202 260-5244
Fax: 202 260-8550
Email: ross.np@epamail.epa.gov
List of speakers who are nonmembers: None

23.5 biometric.soc.05

Sponsoring Section/Society: ENAR

Session Slot: 4:00- 5:50 Sunday
Estimated Audience Size: 125-175
AudioVisual Request: Two Overheads

Session Title: Recent Advances in Nonparametric Methods in Biological/Medical Research

Short description about the session: Talks in this session address the recent developed nonparametric methods with application on Biological/Medical Research. Various approaches including kernel, regression spline and additive modeling will be discussed. Each speaker will use one or more examples to illustrate the application of their approaches on biological or medical problems.

Theme Session: No
Applied Session: Yes

Session Organizer: **Wang, Naisyin** Texas A&M University

Address: Department of Statistics Texas A&M University College Station TX 77843-3143
Phone: 409-845-3141
Fax: (409-845-3144
Email: nwang@stat.tamu.edu

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 5 minutes First Speaker - 30 minutes Second Speaker - 30 minutes Third Speaker - 30 minutes Floor Discusion - 5 minutes

Session Chair: **Wang, Naisyin** Texas A&M University

Address: Department of Statistics Texas A&M University College Station TX 77843-3143
Phone: 409-845-3141
Fax: (409-845-3144
Email: nwang@stat.tamu.edu

1. Functional Data Analysis in Aging Research

Wang, Jane-Ling, University of California, Davis

Address: Division of Statistics University of California, Davis Kerr Hall Davis CA 95616

Phone: 916-752-7624
Fax: 916-752-7099
Email: wnag@wald.ucdavis.edu

Abstract: *Due to the soaring size of the elderly population, there is an increasing interest in aging research both from the public and policy point of view and within the scientific communities. The study of the life history of insects plays a central role in experimental aging research. These experiments are often longitudinal studies in which samples of individuals or samples of cohorts are followed over time. The resulting data are a sample of curves (or functionals). For example, in a study of medflies the reproduction history (in terms of the number of eggs laid per day) of 1,000 female medflies was observed. This yields a sample of 1,000 egg laying curves. In another study, the lifetimes of medflies held in 66 cages (with about 6,000 flies per cage) under different treatment were observed. This gives a survival or hazard function for each cohort. A sample of 66 survival functions or 66 hazard functions is thus observed.*

When a sample of curves or functionals is observed as in the aforementioned examples, one is then faced with the problem of analyzing a sample of curve data. Such data analysis is termed Functional Data Analysis (FDA) and is often nonparametric in nature. It provides an alternative approach to longitudinal data analysis and has been of keen interest recently to many researchers. Methods for analyzing functional data will be discussed. This will include both existing as well as new procedures. The procedures will be illustrated through several biological applications which arise in aging research.

2. Comparing Nonparametric Models

Bowman, Adrian, University of Glasgow

Address: Department of Statistics Mathematics Building University Gardens University of Glasgow GLASGOW G12 8QW United Kingdom
Phone: +44-0141-330-4046
Fax: +44-0141-330-4814

Email: adrian@stats.gla.ac.uk
Bock, Mitchum, University of Glasgow

Abstract: Nonparametric models offer attractive extensions to standard linear models because they are able to provide flexible descriptions of non-linear features which often occur in real data. In particular, the general structure of additive models provides a unifying framework for fitting these. Some biometrical examples will be used to illustrate this. However, the tools for comparing models of this type and drawing inferences on covariate effects are not yet so well developed. Some recent work in this area will be described and illustrated.

3. Nonparametric Regression with Measurement Error

Ruppert, David, Cornell University

Address: School of Operations Research and Industrial Engineering Cornell University Ithaca NY 14853-3801
Phone: 607-255-9136
Fax: 607-255-9129
Email: david@orie.cornell.edu
Carroll, Raymond J., Texas A&M University
Maca, Jeffery D., Cornell University

Abstract: We will describe functional and structural methods for nonparametric regression when the predictor is measured with error. The functional method is the SIMEX estimator of Cook and Stefanski applied to kernel regression: an asymptotic theory is derived. The structural method uses regression splines, with a flexible model for the mismeasured variable.

List of speakers who are nonmembers: None

23.6 biometric.soc.06

Sponsoring Section/Society: ENAR

Session Slot: 8:30-10:20 Thursday
Estimated Audience Size: xx-xxx
AudioVisual Request: xxx

Session Title: Multiple Endpoints and Decision Making in Clinical Trials

The essence of interpretation in a clinical trial is consistency from both the clinical and the statistical perspective. The interpretation of and the conclusions drawn from the results of a clinical trial are dependent on a number of factors that include the disease under study, the patient population, the endpoints, the study design, the conduct of the study, the appropriateness of the statistical analysis for the given design and the sensitivity of the chosen statistical test(s) to the scientific question(s) the study is designed to investigate. One of the many factors that may make such interpretation difficult and sometimes impossible is the presence of multiplicity in a clinical trial that is unaccounted for in the design and ensuing statistical analyses.

One of the incentives for wanting to identify more than one endpoint in a clinical trial is the compelling argument that the conclusion of therapeutic benefit would be strengthened if treatment effectiveness is demonstrated consistently across multiple endpoints. This invited session examines the use of exact, stable, resampling, and nested statistical tests to address the Type I error rate and power issues during the design and analysis of clinical trials with multiple endpoints.

Theme Session: Yes
Applied Session: Yes

Session Organizer: **Sankoh, Abdul** Food and Drug Administration

Address: Division of Biometrics III, HFD-720 FDA/CDER/OEB 5600 Fishers Lane Rockville MD 20857
Phone: 301-827-3090
Fax: 301-443-9279
Email: Sankoha@cder.fda.gov

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 5 First Speaker - 30 minutes Second Speaker - 30 minutes Third Speaker - 30 minutes Discussant - 10 minutes Floor Discusion - 5 minutes

Session Chair: **Sankoh, Abdul** Food and Drug Administration

Address: Division of Biometrics III, HFD-720 FDA/CDER/OEB 5600 Fishers Lane Rockville MD 20857
Phone: 301-827-3090
Fax: 301-443-9279
Email: Sankoha@cder.fda.gov

1. Characterizing the Multivariate Treatment Difference in Clinical Trials by Multiple Nested Tests

Tang, Dei-In, Nathan S. Kline Institute of Psychiatric Research

Address: Statistical Sciences and Epidemiology Division Nathan S. Kline Institute of Psychiatric Research Orangeburg NY 10902
Phone: 914-356-2000
Fax: 914-359-7029
Email: Tang@rfmh.org

Abstract: Clinical Trials comparing two treatments usually use multiple enpoints to measure treatment difference. How to characterize the unerlying ymulitvariate treatment difference in a valid, efficient, and clinically meaningful way has been a subject of long term research interest. Various statistical tests have been proposed that are powerful for their respective inference purposes. However, in most cases the null hypothesis is either too small for ites rejection to be sufficient information for the purpose of treatment selection, or too large to be rejected with a reasonable power. Moreover, it seems that any single criterioon for treatment selection, however carefully designed, may turn out to be too loose or too stringent relative to the outcome of the study. Although a confidence set can be used to make a valid inference under any criterion, currently available confidence sets may not be powerful for specific inference purposes. Between conducting a single test of conducting a whole continuum of tests by using a confidence set lies the possibility of conducting multiple tests. The focus of this talk is on how available tests might be used in combination. The strategy is to formulate a nested sequence of null hypotheses that connect existing formulations so that successive rejections would lead to a sequence of conclusions of increased clinical significance.

2. The Application of Exact Stable Multivariate Tests in Clinical Research

Laeuter, Juergan, Otto-Von-Guericke-Universitaet Magdeburg

Address: Institut fuer Biometrie und Medizinische Informatik Otto-Von-Guericke-Universitaet Magdeburg Leipziger Strasse 44 39120 Magdeburg Germany

Phone: +49 391 6713535

Fax: +49 391 6713536

Email: Juergen.Laeuter@medizin.uni-magdeburg.de

Abstract: In 1995 we proposed a principle of multivariate testing which represents an alternative to the classical least squares mehtod. This principle is based on forming linear combinations of the given p variables, the so-called lienar scores, that are spherically distributed under the null hypothesis of the test. The spherical distribution enables us to perform exact t-tests, F-tests, etc. in the same way as traditional statistics. This strategy of testing is available for cases with a very small sample size n and, nevertheless, with a large number of variables p. A broad range of possibilities for the definition of the score coefficients allow data pre-processing that is suitable to the given application. Thus exploratory statistical methods become an essential part of the confirmatory procedures, not impairing the level of significance alpha of the test.

In this talk, some tests for the comparison of mean values are considered which can be useful in clinical research in situations with multiple endpoints: the standardized sum test (SS test), the principal component test (PC test), and the covariance sum test (CS test), each without and with selection of variables. If the multiple endpoints fulfill corresponding assumptions, these tests have good properties of numerical and statistical stability, i.e. they possess a high power with a large number of variables p and with high correlation between the variables. Furthermore, we will show that the PC test is admissible in the sense of decision theory, i.e. that it cannot be surpassed by another score-based test with respect to power in all cases of application.

Moreover, the tests treated here may be modified by taking into account additional conditions of practical applications: symmetries between the variables, the directions of their effects, some special profiles of their effects, their mutual similarities, their relationship to certain latent variables, the possibility to eliminate redundant variables, and so on.

Based on the theory of spherical distributions, the claissical tests of the repeated measurements analysis can be replaced by certain more general exact multivariate tests which no longer need the assumption of compound symmetry. In these tests, a decomposition of the mean value curve into uncorrelated components corresponding to the treatment and time effects is utilized.

The methods will be demonstrated by some examples from neurological research.

3. Muliple Endpoints in Clinical Trials: A Regulatory Perspective

Jin, Kun, Food and Drug Administration

Address: Division of Biometrics I, HFD-710 Woodmont II FDA/CDER/OEB 5050 Rockville Pike Rockville MD 20857

Phone: 301-594-5302

Fax: 301-594-6593

Email: Jink@cder.fda.gov

Chi, George, Food and Drug Administration

Abstract: Clinical trials often have multiple endpoints, and handling multiple endpoints is a difficult problem in evaulation of the trials. We propose first to define what we mean by primary, co-primary, and secondary endpoints. Then we discuss how clinical decision rules whould be defined in terms of these endpoints. The hypothesis should then be defined to reflect the clinical decision rule with appropriate testing procedure. To design an appropriate testing procedure with accurate Type I error, we must adequately account for correlations among test statistics. We will discuss how to assess correlations of different types of test statistics and how this information will be used in the evaluation of clinical trials.

Discussant: **D'Agostino, Sr., Ralph** Boston University

Address: Department of Mathematics Boston University 111 Cummington Street Boston MA 02215
Phone: 617-353-2767
Fax: 617-353-8100
Email: ralph@math.bu.edu
List of speakers who are nonmembers: None

23.7 biometric.soc.07

Sponsoring Section/Society: WNAR

Session Slot: 10:30-12:20 Monday
Estimated Audience Size: xx-xxx
AudioVisual Request: xxx

Session Title: Statistical Inference for Genetics

Theme Session: No
Applied Session: No

Session Organizer: **Lazzeroni, Laura** Stanford University

Address: Department of Statistics Stanford University Stanford, CA 94305
Phone: (415) 723-0947
Fax: (415) 725-8977
Email: laura@stat.stanford.edu

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 5 or 0 minutes First Speaker - 30 minutes (or 25) Second Speaker - 30 minutes Third Speaker - 30 minutes Discussant - 10 minutes (or none) Floor Discusion - 10 minutes (or 5 or 15)

Session Chair: **Guerra, Rudy** Southern Methodist University

Address: Department of Statistical Science Southern Methodist University 3225 Daniel Avenue Dallas TX 73275-0332
Phone: (214) 768-2770
Fax:
Email: rguerra@mail.smu.edu

1. Evolutionary Graph Methods to Fine-map Disease Genes
Roeder, Kathryn, Carnegie Mellon University

Address: Department of Statistics Carnegie Mellon University Pittsburgh, PA 15213
Phone: 412-268-2717
Fax: 412-268-7828
Email: roeder@stat.cmu.edu

Abstract: Recombinant mapping frequently places the location of a disease locus a few million base pairs. Various methods, all relying on population-level association between the disease locus and adjacent marker-loci, have been proposed to refine or fine-map disease genes. One of these methods uses haplotype data to infer the location of ancestral recombinant breaks. But there are no rigorous analytic techniques underlying this method, in large part because the processes generating the data are complex. We propose analytic methods that rely on graph theory and penalized likelihood to build evolutionary graphs that direct the search for the disease locus. We use evolutionary simulations to demonstrate the strengths and weaknesses of our methods and to draw distinctions between them and the 'by-eye' methods currently used for haplotype fine-mapping.

2. Directed Stratgies for Sequencing the Genome
Lazzeroni, Laura, Stanford University

Address: Division of Biostatistics HRP Redwood Bldg., Stanford, CA 94305-4065
Phone: (415) 723-0947

Fax: (415) 725-8977
Email: laura@stat.stanford.edu

Abstract: *It is impossible to sequence the entire genome or even a single chromosome using current laboratory techniques. In fact, the read length of a typical DNA sequence extends only a few hundred bases. Shotgun sequencing is one strategy used to cover longer regions. In that approach, the genome is randomly fragmented into shorter pieces that are sequenced and then reassembled into larger units. In contrast, directed sequencing uses various information to select fragments before sequencing them in their entirety. Directed sequencing can eliminate much redundant work and yield a more predictable process for completing a given region. This talk will address issues of probabilistic modeling, statistical estimation and optimization encountered in designing efficient and predictable strategies for directed sequencing.*

3. Genetic Linkage Analysis and Change-point Problems

Siegmund, David, Stanford University

Address: Department of Statistics Sequoia Hall, Stanford, CA 94305-4065
Phone: 650-723-0598
Fax: 605-725-8977
Email: dos@stat.stanford.edu

Abstract: *The goal of genetic linkage analysis is to locate genes related to particular traits, e.g., genes that may increase human susceptibility to particular diseases or genes that may increase productivity of agriculturally important species. The availability of increasing numbers of informative genetic markers at known locations makes it possible to study larger numbers of traits than was previously possible, and the design and analysis of the the appropriate experiments raise new statistical questions. I will review the genetic background of modern linkage analysis with special emphasis on three statistical problems: (i) multiple comparisons in the detection of linkage; (ii) the power to detect linkage*

as a function of the true genetic model, sample size, marker density and marker heterozygosity; and (iii) localization of genes by confidence regions. Some of these problems can be understood in terms of recent literature on change-point problems, to which they are closely related. Particular attention will be given to complex and quantitative traits, which one expects to involve multiple, possibly interacting, genes, and to the relative efficiency of different strategies.

Discussant: **Guerra, Rudy** Southern Methodist University

Address: Dept. of Statistics Southern Methodist University Dallas, TX 75275
Phone:
Fax:
Email:
List of speakers who are nonmembers: None

23.8 biometric.soc.08

Sponsoring Section/Society: WNAR

Session Slot: 8:30-10:20 Monday
Estimated Audience Size: xx-xxx
AudioVisual Request: xxx

Session Title: Exciting Developments In Methods For Repeated Measures

Theme Session: No
Applied Session: No

Session Organizer: **Weiss, Robert** Univ. of California at Los Angeles

Address: Department of Biostatistics UCLA School of Public Health Los Angeles, CA 90095-1772
Phone: (310) 206-9626
Fax: (310) 267-2113
Email: rob@rem.ph.ucla.edu

Session Timing: 110 minutes total (Sorry about format):

30 Minutes First Speaker 30 minutes Second Speaker 30 minutes Third Speaker 0 minutes Discussant - none Floor Discusion - 10 minutes

Session Chair: **Weiss, Robert** Univ. of California at Los Angeles

Address: Department of Biostatistics UCLA School of Public Health Los Angeles, CA 90095-1772
Phone: (310) 206-9626
Fax: (310) 267-2113
Email: rob@rem.ph.ucla.edu

1. Functional Linear Models for Longitudinal Data

Fan, Jianqing, University of North Carolina

Address: Deptment of Statistics University of North Carolina Chapel Hill, NC 27599-3260
Phone: 919-962-1358
Fax: 919-962-1279
Email: jfan@stat.unc.edu

Abstract: Functional linear models enhance significantly the flexibility of traditional linear models by allowing coefficients to depend smoothly on time. They include many useful statistical models used for longitudinal data and functional data. In this talk, a simple and powerful approach is introduced to estimate nonparametrically the coefficient functions of the functional linear models. The approach consists of two steps. The first step gives the raw estimate and its standard error of a coefficient function at each time point via using the traditional least-squares technique. In the second, the raw estimates are smoothed by using an existing nonparametric regression technique. Our approach is very flexible that it copes well with various degrees of smoothness of coefficient functions. Some asymptotic results for estimated coefficient functions are established. The applications to a dataset *on CD4 counts and a dataset on nested samples of progesterone curves demonstrate that the proposed approach is indeed powerful in exploiting fine structure for longitudinal data.*

2. Prior Distributions, Bayesian Computation, and Model Selection For Generalized Linear Mixed Models

Ibrahim, Joseph, Harvard University

Address: Department of Biostatistics, Dana Farber Cancer Institute, Mayer 3A.18, Boston MA
Phone: 617-632-3012
Fax:
Email: Joseph_Ibrahim@dfci.harvard.edu

Abstract: *Generalized linear models serve as a useful class of regression models for discrete and continuous data. In applications such a longitudinal studies, observations are typically correlated. The correlation structure in the data is induced by introducing a random effect, leading to the generalized linear mixed model (GLMM). In this talk, we propose a class of informative prior distributions for the model parameters in a GLMM which are useful for variable subset selection. The prior specification is motivated from the notion of the existence of data from similar previous studies, called historical data, which is then quantified into a prior distribution for the current study. We derive theoretical and computational properties of the proposed priors, as well as examine properties of the implied posterior distributions. In addition, we develop novel computational methods for sampling from the posterior distribution of the parameters and computing posterior model probabilities. The computational methods are based on the idea of hierarchical centering (Gelfand et al., 1996), and are shown to be quite efficient for sampling from the posterior distribution for the models considered here. To compute the posterior model probabilities, we derive several key results that lead to efficient calculation, and show that only posterior samples from the full model are needed to estimate the posterior probabilities for all of the possible subset models. Real data sets are used to demonstrate the methodology.*

3. Recent Developments In The Use Of Antedependence Models For Longitudinal Data

Zimmerman, Dale, University of Iowa

Address: 233 Schafer Hall, Dept. of Statisitcs and Actuarial Science, University of Iowa Iowa City, Iowa 52242
Phone: 319-335-0818
Fax:
Email: dzimmer@stat.uiowa.edu

Abstract: Antedependence (AD) models are a useful, though not widely known, class of models for the covariance structure of longitudinal data. Like stationary autoregressive models, AD models allow for serial correlation within subjects but they are more general in the sense that they do not stipulate that the variance is constant over time nor that correlations between measurements equidistant in time are equal. AD models will be briefly reviewed and some recent developments in the use of structured AD models will be given. A new graphical diagnostic for determining the order of the model also will be described.

List of speakers who are nonmembers: None

23.9 biometric.soc.09

Sponsoring Section/Society: WNAR

Session Slot: 2:00- 3:50 Sunday
Estimated Audience Size: xx-xxx
AudioVisual Request: xxx

Session Title: Biostatistical Applications of Multi-Level Models

Theme Session: No
Applied Session: Yes

Session Organizer: **Witte, John** Case Western Reserve University

Address: Department of Epidemiology and Biostatistics 2500 MetroHealth Drive Cleveland, OH 44109-1998
Phone: (216) 778-8523
Fax: (216) 778-3280
Email: witte@darwin.cwru.edu

Session Timing: 110 minutes total (Sorry about format):

First Speaker - 30 minutes Second Speaker - 30 minutes Third Speaker - 30 minutes Discussant - Floor Discusion - 15 minutes

Session Chair: **Witte, John** Case Western Reserve University

Address: Department of Epidemiology and Biostatistics 2500 MetroHealth Drive Cleveland, OH 44109-1998
Phone: (216) 778-8523
Fax: (216) 778-3280
Email: witte@darwin.cwru.edu

1. Getting It Right? Several Analyses Of a Pediatric Pain Repeated Measures Data Set

Weiss, Robert, UCLA Biostatistics

Address: Department of Biostatistics UCLA School of Public Health Los Angeles, CA 90095-1772
Phone: (310) 206-9626
Fax: (310) 267-2113
Email: rob@rem.ph.ucla.edu

Abstract: I discuss statistical modeling issues in the context of 9 different analyses of a repeated measures data set collected to study the effects of pain coping style (CS) and a treatment. Time pressure usually demands that Statisticians analyze data quickly, resulting in the use of models and methods that are simple and easily available in standard software. Occasionally we have time to take one data set thoroughly apart so as to make an attempt at getting the model completely correct, or, in this case, to try out several different analyses. This talk sketches a portion of several *analyses of a Pediatric pain repeated measures data set. The data consists of up to 4 repeated pain tolerance trials on 64 kids. Pain tolerance is assessed by the length of time the children can keep an arm immersed in cold water. Some of the problems with the data set are (a) some missing data, (b) censoring of large responses, (c) repeated measures, (d) non-constant variance, (e) skewness and (f) a lack of overwhelming statistical significance.*

2. Assessment of Gene-Environment Interactions with Hierarchical Models

Aragaki, Corinne, Fred Hutchinson Cancer Research Center

Address: Quantitative Genetic Epidemiology Metropolitan West Building, Room #805 Seattle WA 98104
Phone: 206-667-7256
Fax:
Email: caragaki@cclink.fhcrc.org

Abstract: Assessment of gene-environment interactions with hierarchical models is proposed. The properties of the estimates are studied in finite samples with a limited simulation study. This method is illustrated with an application to diet-NAT2 interaction estimation in a colorectal polyp data set.*

3. Radon Dose Imputation in the Colorado Plateau Miners Chort Using a Hierarchical Model for the Temporal and Geographical Distribution of True Exposure

Stram, Daniel, Univ. of Southern California

Address: Suite 220, 1540 Alcazar Street Los Angeles, CA 90033
Phone: 213-342-1817
Fax: 213-342-2349
Email: stram@rcf.usc.edu
Huberman, Mark, Univ. of Southern California

Abstract: This talk will describe measurement error correction of analyses of the Colorado Uranium Miners Cohort lung cancer data based upon a hierarchical model for the geographic and temporal distribution of radiation dose in the Colorado Plateau from 1950-1969. The implications of the hierarchical model's dose imputations in a realistic model for lung cancer mortality will be addressed.

List of speakers who are nonmembers: None

23.10 biometric.soc.10

Sponsoring Section/Society: WNAR

Session Slot: 2:00- 3:50 Tuesday
Estimated Audience Size: xx-xxx
AudioVisual Request: xxx

Session Title: Advances in Inference Based on Randomization

Theme Session: No
Applied Session: No

Session Organizer: **Zerbe, Gary** University of Colorado

Address: University of Colorado Health Sciences Center Campus Box B119 4200 East 9th Ave Denver, CO 80262
Phone: (303) 315-7608
Fax: (303) 315-3183
Email: Gary.Zerbe@UCHSC.edu

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 0 minutes First Speaker - 30 minutes Second Speaker - 30 minutes Third Speaker - 30 minutes Discussant - 10 minutes Floor Discusion - 10 minutes

Session Chair: **Zerbe, Gary** University of Colorado

Address: University of Colorado Health Sciences Center Campus Box B119 4200 East 9th Ave Denver, CO 80262
Phone: (303) 315-7608
Fax: (303) 315-3183
Email: Gary.Zerbe@UCHSC.edu

1. Comparing Sample Means And Variances

Manly, Bryan, University of Otago

Address: Mathematics and Statistics Department University of Otago, Dunedin, New Zealand
Phone: 64-3-479-7775
Fax: 64-3-479-8427
Email: bmanly@maths.otago.ac.nz
Francis, Chris, National Institute of Water and Atmospheric Research, New Zealand

Abstract: *A common problem is that two or more samples need to be compared in terms of mean and variation differences. What is desirable is a test to compare mean values that allows the samples to be from sources with different amounts of variation, and a test to compare the variation in the samples that allows for the samples to be from sources with different means. Such tests are available with good properties, based on randomization inference, for samples from distributions that are not too far from normal (e.g. normal distributions themself, uniform distributions, and exponential distributions), but their behavior deteriorates badly for the extreme distributions that often occur in biological settings (e.g., with a high proportion of zero values, some moderate values, and a few extremely large values). For these extreme distributions simulation shows that it is sometimes not possible to separate mean differences from variation differences.*

Given this unfortunate situation, one approach is to define a condition number for a set of data based on the sample skewness and kurtosis such that for data with a low number the tests for

means and variation are reliable while for data with a high condition number these tests are not reliable. In that case, for tests with a high condition number it is still possible to test for evidence of some sample differences, without being able to specify exactly the nature of these differences. In a court of law situation, for example, this would then at least be better than not knowing whether the tests on means and variation are reliable or not.

In this talk these ideas will be discussed in terms of the definition of the condition number and the various tests that should be used. The results of a simulation study will be presented to justify the conclusions reached.

2. Theory of Randomization Tests

Edington, Eugene, eedgingt@acs.ucalgary.ca

Address: Department of Psychology University of Calgary Calgary, Alberta T2N 1N4 Canada

Phone: 403-234-9607

Fax: 403-282-8249

Email: eedgingt@acs.ucalgary.ca

Abstract: Randomization tests are permutation tests for testing hypotheses about experimental treatment effects and require experimental randomization (random assignment), whereas permutation tests for testing hypotheses about populations in nonexperimental research are based on random sampling. Despite the differences between the type of hypotheses tested, the same computational procedures and therefore the same computer programs may be used for determining significance for both applications. It is not surprising, therefore, that much of randomization test theory is general permutation test theory that is relevant to both experimental and nonexperimental applications.

The validity of widely differing methods of significance determination, which may yield somewhat different p-values, can be demonstrated by showing that the different reference sets for determining significance each has a property called "stochastic closure." Stochastic closure ensures validity of a reference set

whether that set is produced systematically, randomly, or semi-randomly. That is true of permutation tests based on random sampling as well as randomization tests, which are based on random assignment. The choice of one method over another may be on the basis of which is simpler, faster, or otherwise better for computer programming. Alternatively, preference could be based on relative power or, say, whether a random reference set is the only one possible for the experimental design.

It can be shown that subjects (or other experimental units) may be selected in any systematic (nonrandom) manner to fit the needs of the experimenter without threatening the validity of randomization tests, a fact that is of considerable importance, but equally important is the flexibility permitted in assigning those units. Randomization tests can be developed to accommodate any experimental design in which there is some form of random assignment of experimental units. The development of new experimental designs is thereby facilitated, the value of which has been well illustrated in recent years by the success of single-subject randomized clinical trials.

3. An Empirical Comparison of Permutation Methods for Tests of Partial Regression Coefficients in a Linear Model

Legendre, Pierre, University of Monreal

Address: Department of Biological Sciences University of Montreal C.P. 6128, succ. Centre-Ville Montreal, Quebec H3C 3J7 Canada
Phone: 514-343-7591
Fax: 514-343-2293
Email: legendre@ERE.UMontreal.CA
Anderson, Marti, University of Montreal

Abstract: This study compares empirical type I error and power of different permutation techniques for a test of significance of a single partial regression coefficient in a multiple regression model, using simulations. The methods compared were permutation of raw data values, permutation of residuals under a null (reduced) model and permutation of residuals under the full

model. We investigated effects of changes in 1) sample size, 2) degree of collinearity between the predictor variables, 3) size of the parameter not being tested and 4) distribution of the added random error term. We obtained several replicate sets of simulations so that statistical comparisons of error and power could be made. The type I error was maintained at the chosen level of significance (a = 0.05) for permutation of raw data only when the parameter not being tested was equal to zero. Two methods which had been identified as equivalent formulations of permutation under the null model were found to be different, with one of the formulations giving consistently inflated type I error, while the other gave results with type 1 error close to nominal significance levels. Permutation of residuals under the null model had type I error closer to a than permutation under the full model at very small sample sizes (n = 9). We detected slightly better power using permutation under the full model, compared to permutation under the null model, with increases in the size of the parameter being tested. All methods suffered reductions in power with increases in collinearity of the predictor variables. In general, the use of permutation of raw data for tests of partial regression coefficients cannot be recommended. Either of the model-based permutations had similar behaviour, generally maintaining quite good level accuracy and comparable power.

List of speakers who are nonmembers: Bryan Manly

23.11 biometric.soc.11

Sponsoring Section/Society: WNAR

Session Slot: 10:30-12:20 Thursday
Estimated Audience Size: xx-xxx
AudioVisual Request: xxx

Session Title: Application of Adaptive Cluster Sampling to Biological Populations

Theme Session: No
Applied Session: Yes

Session Content: Public policy is linked to management of biological populations because population changes can have differential consequences to society. For example, society faces tradeoffs when conservation of rare or endangered species interferes with other public and private uses of land, water, or mineral resources. Alternatively, consumers of commercially valuable species, such as ocean fisheries, can have their interest curtailed when population size drops to a point where managers restrict harvest. In either case, accurate estimates of population parameters are critical to successful conservation and informed public debate. However, due to spatial clustering and rarity of many biological populations, population parameters are often difficult to estimate accurately, and yield of individual plants or animals is often low. Application of conventional finite-population sampling schemes to rare and clustered populations may not provide sufficient information to make effective policy given the sampling effort usually afforded to biological surveys.

Adaptive cluster sampling is a class of unconventional sampling techniques specifically designed for sampling rare and clustered populations. It increases sampling effort where high abundance is observed. Its application to study of highly aggregated biological populations promises more information - more precise estimates, higher yield of individuals - compared to non-adaptive sampling techniques (e.g., waterfowl wintering in central Florida, Pacific hake larvae off California coast, freshwater mussels in eastern North American rivers). However, questions still remain about its application: what degree of rarity and clustering is required for adaptive cluster sampling to be efficient, how can design factors be set optimally, how can final sample size be limited? Answers to these questions likely will precede widespread application of adaptive sampling techniques to the study of biological populations.

This session will bring together researchers active in the theoretical development and application of adaptive cluster sampling techniques to present and discuss recent developments, effective application, and important areas of research.

Session Organizer: **Lo, Nancy C.H.** National Marine Fisheries Service

Address: Southwest Fisheries Science Center P.O. Box 271 La Jolla, CA 92038
Phone: 619-546-7123
Fax: 619-546-5656
Email: nlo@ucsd.edu Co-organizer of the session is David R. Smith, U.S.

Geological Survey, Biological Resources Division. He is also the third speaker and contact information for him is provided below.

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 5 minutes First Speaker - 25 minutes Second Speaker - 25 minutes Third Speaker - 25 minutes Discussion lead by Discussant - 20 minutes Floor Discussion lead by Chair - 10 minutes

Session Chair: **Lo, Nancy C.H.** National Marine Fisheries Service

Address: Southwest Fisheries Science Center P.O. Box 271 La Jolla, CA 92038
Phone: 619-546-7123
Fax: 619-546-5656
Email: nlo@ucsd.edu

1. Restricted Adaptive Cluster Sampling

Brown, Jennifer, Department of Mathematics, University of Canterbury

Address: Biomathematics Resarch Centre Department of Mathematics and Statistics University of Canterbury Private Bag 4800, Christchurch New Zealand
Phone: 03-364-2987 ext 7684
Fax: 03-364-2587
Email: J.Brown@math.canterbury.ac.nz

Abstract: Adaptive cluster sampling can be an efficient design for sampling rare and patchy populations. With this design *the initial sample size is fixed but the size of the final sample (and total sampling effort) can not be predicted prior to sampling. For some populations the final sample size can be quite variable depending on the level of patchiness. Restricted adaptive cluster sampling is a proposed modification where a limit is placed on the sample size prior to sampling and quadrats are selected sequentially for the initial sample. As a result there is less variation in the final sample size and the total sampling effort can be predicted with some certainty, which is important for many ecological studies. Estimates of density are biased with the restricted design but under some circumstances the bias can be estimated by bootstrapping.*

2. Sequential Sampling Designs to Estimate Abundance of Rare Populations

Christman, Mary C., American Univeristy

Address: Department of Mathematics and Statistics American University 4400 Massachusetts Ave NW Washington, DC 20016-8050

Phone: 202-885-3123

Fax: 202-885-3155

Email: xman@american.edu

Abstract: Of interest is designing sampling schemes for rare populations where we utilize information about the degree of clustering of the elements being sampled. Some existing designs, such as adaptive cluster sampling, have been developed that exploit that information but which can have some problems associated with them. These include a random sample size and a design which behaves poorly when the population is neither rare nor highly clustered. We consider a sequential sampling design that combines the positive aspects of adaptive sampling and of sequential analysis. Several sampling designs that are based on taking samples sequentially until some predefined stopping condition is met and which incorporates an adaptive sampling component are developed. For the designs considered, estimators and measures of their accuracy are given and their behavior analyzed.

3. Observations of a Practicing Adaptive Sampler: Lessons Learned, Conclusions Drawn, and Questions Remaining

Smith, David R., Biological Resources Division, U.S. Geological Survey

Address: U.S. Geological Survey Biological Resources Division 1700 Leetown Rd. Kearneysville, WV 25430
Phone: 304-725-8461x383
Fax: 304-728-6509
Email: david_r_smith@usgs.gov

Abstract: Adaptive cluster sampling methods are intuitively appealing because they permit biologists to do what comes naturally - increase sampling effort where rare organisms are discovered. However, it is my impression that while biologists are intrigued by the technique, widespread application is lagging. Through a review of the design of a biological survey that makes use of adaptive cluster sampling, I will highlight the decisions that give biologists pause. Much of the problem in design of such a survey is avoiding over and under sampling. I will illustrate potential problems and make some recommendations based on how I have practiced adaptive cluster sampling. Finally, I will question whether increased efficiency is sufficient justification for application of adaptive cluster sampling, especially for rare populations; rather I will suggest that increased yield of individuals and species is its main benefit.

Discussant: **Thompson, Steve K.** Pennsylvania State University

Address: Department of Statistics Pennsylvania State University 326 Classroom Bldg. University Park, PA 16802-2111
Phone: 814-865-3235
Fax: 814-863-7114
Email: skt@stat.psu.edu

List of speakers who are nonmembers: None ????

23.12 biometric.soc.12

Sponsoring Section/Society: WNAR

Session Title: Managing Natural Resources In The Face Of Uncertainty

Session Slot: 8:30-10:20 Wednesday
Estimated Audience Size: xx-xxx
AudioVisual Request: xxx
Theme Session: Yes
Applied Session: Yes

Session Organizer: **Kendall, William** Patuxent Wildlife Research Center

Address: U. S. Geological Survey, 11510 American Holly Dr., Laurel, MD 20708-4017
Phone: 301/497-5868
Fax: 301/497-5666
Email: William_Kendall@usgs.gov

Session Timing: 110 minutes total (Sorry about format):

> Opening Remarks by Chair - 0 minutes First Speaker - 25 minutes Second Speaker - 25 minutes Third Speaker - 25 minutes Fourth Speaker - 25 minutes (or none) Floor Discusion - 10 minutes

Session Chair: **Kendall, William** Patuxent Wildlife Research Center

Address: U. S. Geological Survey, 11510 American Holly Dr., Laurel, MD 20708-4017
Phone: 301/497-5868
Fax: 301/497-5666
Email: William_Kendall@usgs.gov

1. Data, Policy, And The Conservation Of Renewable Natural Resources

Williams, Byron, U. S. Geological Survey

Address: Cooperative Research Division, U. S. Geological Survey,Biological Resources Division, Reston, VA.
Phone: 703/358-1784
Fax: 703/358-2282
Email: Ken_Williams@fws.gov

Abstract: The conservation of renewable natural resources involves careful monitoring and assessment, as well as strategic planning and on-the-ground conservation. Ideally, conservation is guided through strategic planning, with planning itself informed by monitoring and assessment. In this scenario data collection and analysis are neither ancillary nor posterior to conservation, but instead lie at the heart of the enterprise, with monitoring and assessment both informing, and being informed by, conservation actions. This delivery framework describes an iterative cycle that potentially leads to improved conservation actions over time. Technical elements of the framework include conservation objectives, potential conservation strategies, system responses to management, measures of risk/benefit, and monitoring programs. Policies for conservation arise from the interplay of these elements, pursuant the identification of optimal or satisficing conservation actions. Conservation and management of renewable natural resources differ from other managed systems in the degree of stochasticity, lack of knowledge about biological processes, limitations in the availabilility of data, and the inability to agree upon goals, objectives, and acceptable management strategies. A number of impediments to rational policy development contribute to these problems, including a lack of long-term institutional commitment, limited fiscal and other resources, lack of understanding of system responses to management, political considerations. Some creative approaches to overcome these limitations are illustrated with examples of waterfowl harvest management, fisheries management, and forest management.

2. Environmental Decision Making Under Uncertainty: Tools, Techniques, And Examples

Lubow, Bruce, Colorado Cooperative Fish and Wildlife Research Unit

Address: Colorado State University, 1912 Bronson Street, Fort Collins, CO 80526
Phone: 970/229-9637
Fax: 970/491-1413
Email: bruce@cnr.colostate.edu

Abstract: The paradigm of Markovian decision analysis provides the framework for making decisions under uncertainty. This paradigm includes optimal dynamic control of deterministic and stochastic systems as well as adaptive learning in imperfectly known systems. These techniques are broadly applicable to diverse fields in engineering, operations research, medicine, biology, and natural resource management. I briefly introduce the requirements for posing problems within this conceptual framework and the advantages and limitation of the approach. Next, I present examples of applications to natural resource management and conservation biology that illustrate the flexibility and broad applicability of the technique. I focus on the practical aspects involved in conducting these analyses and the tools and techniques available to assist in this process. I discuss a software package (SDP/ASDP) that solves a wide range of Markovian optimization problems including stochastic and Bayesian adaptive optimization cases. Finally, I present some statistics that provide insight into the current computational limitations on the methodology and describe current techniques and future directions for addressing this fundamental obstacle.

3. Ecological Variation And The Question Of Scale In Natural Resource Management

Johnson, Fred, U. S. Fish and Wildlife Service

Address: Office of Migratory Bird Management 11500 American Holly Dr., Laurel, MD 20708-4016
Phone: 301/497-5861

Fax: 301/497-5871
Email: Fred_A_Johnson@fws.gov

Abstract: *All ecological systems exhibit variability on a broad range of temporal, spatial, and organizational scales as a function ultimately of how individuals respond to their environment. The scale at which individuals are aggregated for management purposes is an arbitrary decision, but one that can strongly influence both the benefits and costs of management. Management systems defined at scales that result in relatively low ecological variation will produce relatively high benefits, but also are characterized by high monitoring and assessment costs. Determining the optimal scale for management depends critically on the availability of explicit performance criteria and on descriptions of relevant ecological patterns. The description of ecological patterns, in turn, involves the use of data to explore variation as a function of scale and to elucidate underlying causal mechanisms. I illustrate these concepts of variation and scale using the management of waterfowl harvests in the United States as an example. I discuss the gradual change of management scale from one of high heterogeneity to one of lower heterogeneity, and provide some thoughts about the increase in net benefit that might be expected.*

4. Data And Wildlife Management In Missouri

Sheriff, Steven, Missouri Department of Conservation

Address: 1110 S College Ave., Columbia, Missouri 65201
Phone: 573/882-9880
Fax: (573)882-4517
Email: Sheris@mail.conservation.state.mo.us

Abstract: *In the United States, individual states are responsible for the management of wildlife resources within their borders. Each state has an agency that is responsible for the management, protection and welfare of their wildlife resources. Citizens are allowed to use these resources in a number of ways including aesthetic viewing, hunting, and general sense of natural well-being.*

To manage their wildlife resources, state agencies use a wide-varity of data in the decision process. For this presentation, I will focus on methods of data collection and types of data that the Missouri Department of Conservation uses in managing our wildlife resources. Examples that will be highlighted include a large scale experiment that is examining the impacts of forest management on components of the forest ecosystem, radio-telemetry studies of wild turkey population dynamics, and attitude surveys of public and hunter constituency groups. How these data are incorporated into the decision process will also be addressed.

List of speakers who are nonmembers: None

Chapter 24

Institute of Mathematical Statistics (24)

Program Chair: **Lindsay, Bruce G.**

24.1 ims.01

IMS

Session Slot: 2:00- 3:50 Wednesday
Estimated Audience Size:
AudioVisual Request:

Session Title: Advances in Hypothesis Testing

Theme Session: No
Applied Session: Yes

Session Organizer: **Brown, Larry** University of Pennsylvania

Address: Statistics Department, University of Pennsylvania, Philadelphia, PA, 19104
Phone: 215-898-4753
Fax: 215-898-1280
Email: lbrown@compstat.wharton.upenn.edu

Session Timing: 110 minutes total (Sorry about format):

Session Chair: **Brown, Larry** University of Pennsylvania

Address: Statistics Department, University of Pennsylvania, Philadelphia, PA, 19104
Phone: 215-898-4753
Fax: 215-898-1280
Email: lbrown@compstat.wharton.upenn.edu

1. On Testing of Individual Bioequivalence

Wang, Weizhen, Wright State University

Address:
Phone:
Fax:
Email: wwang@euler.math.wright.edu

Abstract: The hypothesis of individual bioequivalence typically involves several parameters, such as the mean and variance of two populations: the generic drug and the brand name drug. The hypothesis spaces may be complicated regions in a plane or even in a 3-dimensional space. How can we construct tests involving these kinds of hypothesis regions when the data are normally distributed? In this talk, a reparametrization is introduced through a linear regression approach. A general class of hypotheses, which includes individual bioequivalence, is discussed and the exact alpha-level tests are proposed. When there is no interaction between formulation and subject, a two-by-two crossover design is sufficient to assess individual bioequivalence, while a two-by-three crossover design should be used if interaction is present.

Key words: Linear regression; Noncentral t-distribution; Power; Reparametrization.

2. Mode Testing in Difficult Cases

Cheng, Ming-Yen, National ChungCheng University, (Taiwan)

Address: Institute of Mathematical Statistics, National Chung Cheng University, Taiwan

Phone:

Fax:

Email: mycheng@math.ccu.edu.tw

Hall, Peter, Australian National University, Australia

Address: Centre for Mathematics and its Applications, Australian National University, Australia

Abstract: Usually, when testing the null hypothesis that a distribution has one mode, against the alternative that it has two, the null hypothesis is interpreted as entailing that the density of the distribution has a unique point of zero slope, which is a local maximum. We argue that a more appropriate null hypothesis is that the density has two points of zero slope, of which one is a local maximum and the other is a shoulder. The choice of null hypothesis is important when calibrating the test, and there are at least two reasons for favouring the mode-with-shoulder form. First, a unimodal density with a shoulder is the more difficult case to distinguish from bimodality, and in general a test should be constructed in such a way that it works best in difficult cases. (Here, we use the word "difficult" to refer to the difficulty of distinguishing between null and alternative hypotheses.) Secondly, when a test for a mode-with-shoulder is properly calibrated, so that it has asymptotically correct level, it is generally conservative when applied to the case of a mode-without-shoulder. Doing things the other way around would produce a liberal, rather than conservative, test in the important, difficult case of a mode with shoulder. The calibration method, for the bandwidth and excess mass tests, that we suggest involves the bootstrap. In contrast to other approaches it has very good adaptivity properties.

3. Optimal nonparametric testing of qualitative assumptions such as monotonicity or convexity

Duembgen, Lutz, Medical University of Luebeck (Germany)

Address: Medizinische Universitat zu Luebeck, Institut fur Mathematik
Phone:
Fax:
Email: duembgen@bilbo.math.mu-luebeck.de

Abstract: Suppose one observes a process Y on the unit interval, where $dV = f_o + n^{-1/2}dW$ with an unknown function parameter f_o, given scale parameter $n \geq 1$ (sample size) and standard Brownian motion W. We propose two types of tests of qualitative nonparametric hypotheses about f_o such as monotonicity or convexity. These tests are asymptotically optimal and adaptive with respect to two different criteria. As a by-product we obtain an extension of Lévy's modulus of continuity of Brownian motion which is of independent interest. It has potential applications to simultaneous confidence intervals in nonparametric curve estimation. (L. Duembgen and V. G. Spokoiny)

List of speakers who are nonmembers:

24.2 ims.02

IMS

Session Slot: 2:00- 3:50 Monday
Estimated Audience Size:
AudioVisual Request: Two Overheads

Session Title: Neyman Lecture

Theme Session: No
Applied Session: Yes

Session Organizer: **Lindsay, Bruce** The Pennsylvania State University

Address: 422 Thomas Building, Department of Statistics, University Park, PA16802

Phone: (814)865-1220
Fax: (814) 863-7114
Email: bgl@psu.edu

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 5 First Speaker - 90 minutes Floor Discussion - 10 minutes

Session Chair: **Markatou, Marianthi** Columbia University

Address: Department of Statistics, 615 Mathematics Building, New York, NY 10027
Phone: (212) 854 3969
Fax:
Email: markat@stat.columbia.edu

1. Computing with Data: Concepts and Challenges

Chambers, John, Bell Labs, Lucent Technologies

Address: Rm. 2C-282 Bell Laboratories Lucent Technologies 700 Mountain Av. Murray Hill, NJ 07974-0636
Phone: 908 582-2681
Fax: 908 582-3340
Email: jmc@bell-labs.com

Abstract: Our use of computing, as statisticians or members of related professions, takes place in a broader context of activities, in which data is acquired, managed, and processed for any of a great variety of purposes. The underlying purposes are diverse: scientific experimentation, business transactions, or managing society itself (the original meaning of statistics). The activities, of which traditional statistical computing is an important but quantitatively small part, form what I call computing with data.

This talk examines how work in statistical software and related areas can contribute to the activities, and in particular, to science. Such a focus is an appropriate topic for the Neyman lecture, since

our ability to compute with data is often the limiting factor in applying statistical techniques to science today.

Discussion of what has been done, and is being done now, in statistical software is presented around three basic concepts: language*, the ability to express our ideas;* objects*, the structure we can impose on data; and* interfaces*, the means of communication within software and between software and its users. To complement the discussion of concepts, we need to look at the challenges to be faced. Here, one challenge subsumes many of the most important ones: the* scope *of our participation in the larger context. Simply put, we need to greatly enlarge the scope of our activities and of our model for our own rôle in computing with data.*

List of speakers who are nonmembers: None

24.3 ims.03

IMS

Session Slot: 8:30-10:20 Wednesday
Estimated Audience Size:
AudioVisual Request: Two Overheads

Session Title: Inference from Small Samples

Theme Session: No
Applied Session: No

Session Organizer: **Davison, A. C.** Swiss Federal Institute of Technology, Lausanne

Address: Department of Mathematics
Swiss Federal Institute of Technology
1015 Lausanne
Switzerland
Phone: (41)–(0)21–693–5502
Fax: (41)–(0)21–693–4250
Email: davison@dma.epfl.ch

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 5 minutes First Speaker - 25 minutes Second Speaker - 25 minutes Third Speaker - 25 minutes Discussant - 10 minutes Floor Discussion - 10 minutes

Session Chair: **Wang, Soujin** Texas A&M University

Address: Department of Statistics
Texas A&M University
College Station, TX 77843
Phone: 409–845–3164
Fax: 409–845–3144
Email: sjwang@stat.tamu.edu

1. Adjustments to the Likelihood Ratio Statistic

Severini, Thomas A., Northwestern University

Address: Department of Statistics
Northwestern University
2006 Sheridan Road
Evanston IL 60208
USA
Phone: (1)–847–467–1254
Fax: (1)–847–491–4939
Email: severini@nwu.edu

Abstract: Consider a model parameterized by a scalar parameter of interest as well as a nuisance parameter. Inference about the parameter of interest may be based on R, the signed square root of the likelihood ratio statistic. The distribution of R may be approximated, to first order, by a standard normal distribution. To improve the accuracy of this normal approximation, several modifications to R have been proposed. In this talk, I will discuss the properties of Barndorff-Nielsen's modified directed likelihood statistic, R^*, along with approximations to R^* that are more convenient to use in practice. In particular, I will present an approximation to R^* that can be calculated numerically for a wide range of models. The results will be illustrated on several examples.

2. Technology Transfer: Implementing Small-Sample Methods

Davison, Anthony C., Swiss Federal Institute of Technology, Lausanne

Address: Department of Mathematics
Swiss Federal Institute of Technology
1015 Lausanne
Switzerland
Phone: (41)–(0)21–693–5596
Fax: (41)–(0)21–693–4250
Email: Anthony.Davison@epfl.ch

Abstract: Over the past two decades there has been rapid theoretical progress on small-sample asymptotics for parametric statistical models, and although the theory remains incomplete many aspects are now well-understood. However, compared to competing technology such as the bootstrap and Markov chain Monte Carlo methods, small-sample parametric theory seems to have been under-used in applications. One reason for this may be the sometimes forbidding mathematics used, but probably more important is the lack of software with which to apply the parametric approximations in routine data analysis.

In this talk I shall briefly outline what seems to me to be the state of the art in implementation of nonparametric small-sample inference, with special reference to a `S-Plus` *library of bootstrap functions written by A. J. Canty. Then I shall describe the initial phase of work on implementing small-sample parametric methods, which is being performed by A. R. Brazzale, with the goal of making use of* r^* *and related approximations as straightforward as the use of nonparametric methods, in order that wider numerical experience of parametric methods can be gained by the developers and consumers of small-sample asymptotics.*

3. Exact Inference for Sparse Contingency Tables Using Markov Chain Monte Carlo Methods

Smith, Peter W. F., University of Southampton

Address: Department of Social Statistics
University of Southampton
Highfield
Southampton SO17 1BJ
UK
Phone: (44)–(0)1703–593191
Fax: (44)–(0)1703–593846
Email: pws@soton.ac.uk

Abstract: *When analysing sparse contingency tables inferences based on asymptotic distributions may be unreliable, especially for goodness-of-fit tests, and inferences based on exact distributions are preferred. I will present, for a log-linear model, the form of the exact conditional distribution of a sufficient statistic for the interest parameters, given the sufficient statistic for the nuisance parameters. In general, direct generation from this distribution is infeasible. I will describe Markov chain Monte Carlo methods for generating from the conditional distribution. These methods enable estimation of the exact conditional p-value and the exact distribution of the residuals for log-linear models. Examples will include tests of goodness of fit of the all-two-way interaction model, the quasi-symmetry model and the uniform association model. Tests against non-saturated alternatives will also be considered. I will conclude by briefly discussing irreducibility of the Markov chains and other methods for generating from the exact conditional distribution.*

Discussant: **Davison, A. C.** Swiss Federal Institute of Technology, Lausanne

Address: Department of Mathematics
Swiss Federal Institute of Technology
1015 Lausanne
Switzerland
Phone: (41)–(0)21–693–5502
Fax: (41)–(0)21–693–4250
Email: davison@dma.epfl.ch

List of speakers who are nonmembers: A. R. Brazzale, P. W. F. Smith

24.4 ims.04

IMS

Session Slot: 4:00- 5:50 Tuesday
Estimated Audience Size: xxx
AudioVisual Request: Two Overheads

Session Title: President's Invited Session

Theme Session: No
Applied Session: No

Session Organizer: **Diaconis, Persi** Cornell University

Address: Department of Mathematics, ORIE, Cornell University
Phone:
Fax:
Email: persi@orie.cornell.edu

Session Timing: 110 minutes total (Sorry about format):

usual

Session Chair: **Diaconis, Persi** Cornell University

Address: Department of Mathematics, ORIE, Cornell University
Phone:
Fax:
Email: persi@orie.cornell.edu

1. ***Nonparametrics: The Development of a Subdiscipline***

Lehmann, Erich, University of California, Berkeley

Address: Statistics Department, University of California, Berkeley
Phone:
Fax:
Email:

Abstract: *The history of nonparametric inference is used to illustrate how a subdiscipline develops from isolated beginnings, both conceptually and through its dissemination via survey papers, lecture courses, and books.*

2. *Hydrodynamical Scaling and Large Deviations*
Varadhan, S. R. S., New York University

Address: Mathematics Department, New York University
Phone:
Fax:
Email:

Abstract: *In general in probability theory, laws of large numbers are proved first, followed by fluctuation theory or central limit theorems and then the large deviation results are established. However in problems of hydrodynamical scaling, establishing even the laws of large numbers require a fair amount of ideas from large deviation theory. The lecture will consist of a description of what hydrodynamical scaling results are and why large deviation theory is important for proving them.*

3. *Applications of a Heuristic Concerning Partial Asymptotic Equivalence for Multivariate Estimation Problems*
Brown, Lawrence D., University of Pennsylvania

Address: Statistics Department, University of Pennsylvania, Philadelphia, PA 19104-6302
Phone:
Fax:
Email: lbrown@stat.wharton.upenn.edu

Abstract: *Nonparametric function estimation problems include those usually described as nonparametric regression, density estimation, and signal processing. One dimensional versions of these problems have been proved under mild conditions to be*

asymptotically equivalent in a strong sense by Brown and Low and by Nussbaum (both in Ann. Statist., Dec. 1996). Consequently (under mild conditions), to any procedure in one setting there exist corresponding procedures in the others which have identical asymptotic properties.

Multivariate versions of these problems present additional obstacles related to the curse of dimensionality. Consequently, multivariate extensions of the strong equivalence results mentioned above often require unpalatably strong smoothness conditions. (See Brown and Zhang, Ann. Statist., Feb. 1998.) In the present talk we will describe ways to use heuristics from the proof of the strong equivalence theorems in order to suggest weaker results about equivalence. These weaker results relate the asymptotic rates of convergence of appropriate procedures in the respective problems. Even in multivariate situations such partial results require only very mild conditions.

These ideas will be applied in the setting of multivariate tensor product models to derive some familiar and some new ratewise optimality results for appropriate smoothing spline estimators.

List of speakers who are nonmembers:

24.5 ims.05

IMS

Session Slot: 2:00-3:50 Sunday
Estimated Audience Size: 125-175
AudioVisual Request: Two Overheads

Session Title: Model Uncertainty

Theme Session: No
Applied Session: No

Session Organizer: **Draper, David** University of Bath, UK

Address: Department of Mathematical Sciences University of Bath Claverton Down, Bath BA2 7AY England
Phone: +44-1225-826222
Fax: +44-1225-826492
Email: d.draper@maths.bath.ac.uk

Session Timing: 110 minutes total (Sorry about format):

> Opening Remarks by Chair - 0 minutes First Speaker - 30 minutes Second Speaker - 30 minutes Third Speaker - 30 minutes Discussant - (none) Floor Discusion - 15 minutes

Session Chair: **Draper, David** University of Bath, UK

Address: Department of Mathematical Sciences University of Bath Claverton Down, Bath BA2 7AY England
Phone: +44-1225-826-222
Fax: +44-1225-826-492
Email: David Draper ¡d.draper@maths.bath.ac.uk¿

1. A Bayesian decision-theoretic approach to model selection

Walker, Stephen, Imperial College, London

Address: Department of Mathematics Imperial College of Science, Technology, and Medicine Huxley Building 180 Queen's Gate London SW7 2BZ England
Phone: +44-171-594-8522
Fax: +44-171-594-5817
Email: S G Walker ¡s.walker@ic.ac.uk¿

Abstract: *We put the problem of model selection into a Bayesian decision-theoretic framework. Typically, the models under consideration consist of a finite number of parametric models. However, the decision space might also include a nonparametric model. We therefore consider the problem of deciding when to adopt a nonparametric approach in preference to "sticking" with a parametric model.*

2. *Bayesian model averaging*

Madigan, David, University of Washington

Address: Department of Statistics, GN-22 University of Washington Seattle WA 98195-0001 USA
Phone: 206-543-4537
Fax: 206-685-7419
Email: David Madigan ¡madigan@stat.washington.edu

Abstract: *Standard statistical practice ignores model uncertainty. Data analysts typically select a model from some class of models and then proceed as if the selected model generated the data. This approach ignores the uncertainty in model selection, leading to over-confident inferences.*

Bayesian model averaging (BMA) provides a coherent mechanism for accounting for this model uncertainty. Several methods for implementing BMA have recently emerged. We discuss these methods and present a number of examples. We emphasize applications involving latent variables, missing data, and causal inference, where between-model uncertainty can overwhelm within-model uncertainty.

3. *Model-based inference for categorical survey data*

Forster, Jon, University of Southhampton, UK

Address: Department of Mathematics University of Southampton Southampton SO17 1BJ England
Phone: +44-1703-595-130
Fax: +44-1703-595-147
Email: Jon Forster ¡jjf@maths.soton.ac.uk¿

Abstract: *For survey data, model uncertainty may relate both to the underlying "data model" and to models for "non-sampling errors" such as nonresponse bias. In this talk, I consider categorical data and show how different approaches may be used to account for such uncertainty in order to reliably estimate quantities of interest.*

24.6 ims.06

IMS

Session Slot: 8:30-10:20 Monday
Estimated Audience Size:
AudioVisual Request:

Session Title: Stochastic Analysis

Theme Session: No
Applied Session: No

Session Organizer: **Etheridge, Alison** University of Oxford, UK

Address: Department of Statistics, 1 South Parks Road, Oxford, OX1 3TG, UK
Phone: +44 1865 272 860
Fax: +44 1865 272 595
Email: etheridge@stats.ox.ac.uk

Session Timing: 110 minutes total (Sorry about format):

Session Chair: **Adler, Robert** Technion

Address:
Phone:
Fax:
Email:

1. The Ancestral Selection Graph
Krone, Steve, University of Idaho

Address: Dept. of Mathematics, University of Idaho, Moscow, ID 83844
Phone: 208-885-6317
Fax: 208-885-5843
Email: krone@uidaho.edu

Abstract: *Coalescent theory is currently one of the most active areas of population genetics. The idea is to study the genealogy of a (sample from a) population by viewing the history of a gene in the population backwards in time. In the neutral case (i.e., when there is no natural selection in the model), this approach has lead to a dramatic surge in our understanding of population genetics over the past 20 years. In this talk, we show how to use some simple ideas from interacting particle systems to construct genealogies for a large class of models involving selection and mutation. The genealogy of the sample is embedded in a graph which we call the ancestral selection graph. This graph contains all the information about the genealogy of the sample; it is the analogue of Kingman's coalescent process which arises in the case with no selection. The ancestral selection graph can be easily simulated and we outline an algorithm for simulating samples. The main goal is to analyze the ancestral selection graph and to compare it to Kingman's coalescent process. As applications, we will discuss things like the time to the most recent common ancestor of the sample and the probability of identity by descent. (This is joint work with Claudia Neuhauser.)*

2. Stochastic Coalescent

Limic, Vlada, University of California at Berkeley

Address: Department of Statistics #3860 University of California, Berkeley Berkeley, CA 94720-3860
Phone: 510-642-2781
Fax: 510-642-7892
Email: vlada@stat.berkeley.edu

Abstract: *The* Marcus-Lushnikov(M-L) process *was a very popular model of coagulation studied by theoretical physicists in the 1980's. A verbal description is: at time* 0 *start with a configuration of* n *particles, each of mass* 1*, and let it evolve in continuous time according to the rule: each pair of clusters with masses* x *and* y *merges into a single cluster of mass* $x+y$ *at rate* $K(x,y)/n$*. We give a motivation for the model and describe some typical questions in the field. There are many open problems.*

A generalization of the M-L process is the stochastic coalescent $\mathbf{X}(\cdot)$, *a vector-valued Markov process representing merging of possibly infinitely many clusters of mass. The transitions are induced by the dynamics: each pair of clusters with masses x and y merges into a cluster of mass $x + y$ at rate $K(x, y)$. If the rate kernel $K(\cdot, \cdot)$ is constant, this process is well-known* Kingman's coalescent. *Nice probabilistic representations exist also for the additive ($K(x, y) = x + y$) and the multiplicative ($K(x, y) = xy$) coalescent.*

We introduce the entrance boundary problem, and present the solution for the multiplicative case, a joined work with David Aldous.

3. ***Genealogies of Spatially Structured Populations***
Williams, David, Oxford University

Address: Department of Statistics,Oxford University 1 South Parks Road,Oxford OX1 3TG, England
Phone: +44-1865-2 72870
Fax: +44-1865-272595
Email: drew@stats.ox.ac.uk

Abstract: The coalescent process is used to identify useful biological quantities under very simple conditions. We incorporate a spatially dependent offspring law into the background population and compare the effects of such a modification relative to other, spatially structured, coalescent approaches.

List of speakers who are nonmembers:

24.7 ims.07

IMS

Session Slot: 10:30-12:20 Monday
Estimated Audience Size:

AudioVisual Request: Two Overheads

Session Title: First Wald Lecture

Theme Session: No
Applied Session: No

Session Organizer: **Lindsay, Bruce** The Pennsylvania State University

Address: 422 Thomas Building, Department of Statistics, University Park, PA16802
Phone: (814)865-1220
Fax: (814) 863-7114
Email: bgl@psu.edu

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 5 First Speaker - 90 Floor Discussion - 10 minutes

Session Chair: **Diaconis, Persi** Cornell University

Address: Department of Mathematics, ORIE, Cornell University
Phone:
Fax:
Email: persi@orie.cornell.edu

1. Consistency of Bayes Estimators

Freedman, David, University of California at Berkeley

Address: Department of Statistics University of California, Berkeley, CA 94720-4735
Phone: 510-642-2781
Fax:
Email: freedman@stat.berkeley.edu

Abstract: *Consider a finite sequence of independent, identically distributed observations, with a distribution that depends smoothly on a parameter θ. If θ is finite-dimensional, Bayes estimates are consistent in the frequentist sense: as the number of observations grows, the posterior distribution piles up near θ—for almost all sample sequences governed by θ. (Mild additional regularity conditions are needed.) As is often said, "the data swamp the prior." On the other hand, if θ is infinite-dimensional, consistency is the exception not the rule: the prior often swamps the data. I will review some recent work in this area, including Bayesian non-parametric regression and the Bernstein-von Mises theorem for hierarchical linear models.*

List of speakers who are nonmembers: None

24.8 ims.08

IMS

Session Title: Wald Lecture II

Session Slot: 10:30-12:20 Tuesday
Estimated Audience Size:
AudioVisual Request: Two Overheads

Session Title: Wald Lecture II

Theme Session: No
Applied Session: No

Session Organizer: **Lindsay, Bruce** The Pennsylvania University

Address: 422 Thomas Building, Department of Statistics, University Park, PA16802
Phone: (814)865-1220
Fax: (814) 863-7114
Email: bgl@psu.edu

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 5 minutes First Speaker - 90 minutes Floor Discusion - 10 minutes

Session Chair: **Varadhan, S. R. S.** New York University

Address: Mathematics Department, New York University
Phone:
Fax:
Email: annals@cims.nyu.edu

1. *(*From Association to Causation Through Regression)
Freedman, David, University of California at Berkeley

Address: Department of Statistics University of California, Berkeley, CA 94720-4735
Phone: 510-642-2781
Fax:
Email: freedman@stat.berkeley.edu

Abstract: *For nearly a century, investigators in the social sciences have used regression models to deduce cause-and-effect relationships from patterns of association. Path models and automated search procedures are more recent developments. In my view, this enterprise has not been successful. The models tend to neglect the difficulties in establishing causal relations, and the mathematical complexities tend to obscure rather than clarify the assumptions on which the analysis is based.*

Formal statistical inference is, by its nature, conditional. If maintained hypotheses $A, B, C, \ldots$ *hold, then* H *can be tested against the data. However, if* $A, B, C, \ldots$ *remain in doubt, so must inferences about* H*. Careful scrutiny of maintained hypotheses should therefore be a critical part of empirical work—a principle honored more often in the breach than the observance.*

I will discuss modeling techniques that seem to convert association into causation. The object is to clarify the differences among the various uses of regression, and the difficulties in making causal inferences by modeling.

List of speakers who are nonmembers: None

24.9 ims.09

IMS

Session Slot: 10:30-12:20 Wednesday
Estimated Audience Size:
AudioVisual Request: Two Overheads

Session Title: Wald Lecture III

Theme Session: No
Applied Session: No

Session Organizer: **Lindsay, Bruce** The Pennsylvania State University

Address: 422 Thomas Building, Department of Statistics, University Park, PA16802
Phone: (814)865-1220
Fax: (814) 863-7114
Email: bgl@psu.edu

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 5 First Speaker - 90 minutes
Floor Discussion - 10 minutes

Session Chair: **Owen, Art** Stanford University

Address: Department of Statistics Stanford University Stanford, CA 94305
Phone:
Fax:
Email: art@stat.Stanford.EDU

1. De Finetti's Theorem
Freedman, David, University of California at Berkeley

Address: Department of Statistics University of California, Berkeley, CA 94720-4735
Phone: 510-642-2781
Fax:
Email: freedman@stat.berkeley.edu

Abstract: Consider an infinite sequence of independent, identically distributed (IID) random variables; for simplicity, suppose each variable takes only finitely many values. Plainly, the variables are "exchangeable," in the sense that permuting them does not affect their joint distribution. (Multiplication is, after all, commutative.) Consider next a Bayesian statistician whose predictive distribution—in advance of data collection—can be described as follows: (i) a marginal distribution is chosen at random, (ii) a sequence of independent random draws are made from this marginal. Such a predictive distribution is called a "mixture of IID processes." And any such mixture must also be exchangeable. (The expected value of a constant is constant.) A celebrated theorem of de Finetti's asserts the converse: any exchangeable distribution is a mixture of IID processes.

De Finetti's theorem has been generalized in many different directions. For example, the state space can be the real line (or any complete, separable metric space). Mixtures of specific parametric families of distributions can be characterized. So can mixtures of Markov chains. Bounds can be given for finite sequences of random variables. I will review some of the work on de Finetti's theorem.

List of speakers who are nonmembers: None

24.10 ims.10

IMS

Session Slot: 2:00- 3:50 Sunday
Estimated Audience Size:

AudioVisual Request:

Session Title: Exact Simulation

Theme Session: No
Applied Session: No

Session Organizer: **Green, Peter** University of Bristol

Address: Department of Mathematics, University of Bristol, Bristol, BS8 1TW, UK.
Phone: +44-117-928-7967
Fax: +44-117-928-7999
Email: P.J.Green@bristol.ac.uk

Session Timing: 110 minutes total (Sorry about format):

Session Chair: **Liu, Jun** Stanford University

Address:
Phone:
Fax:
Email:

1. Perfect and Imperfect Simulation

Tweedie, Richard, Colorado State University

Address: Department of Statistics, Colorado State University, Fort Collins CO 80523, USA
Phone: (970) 491 6786
Fax: (970) 491 7895
Email: tweedie@stat.colostate.edu

Abstract: We describe algorithms for simulating random samples from the invariant measures of stochastically ordered Markov chains on spaces which may be continuous. Both "vertical" and

"horizontal" backwards coupling algorithms are introduced. Convergence properties of these algorithms are described, and methods of bounding the results when there are no maximal or minimal elements are outlined.

2. Exact Sampling for Bayesian Inference: Towards General Purpose Algorithms

Murdoch, Duncan, Queens University, Ontario, Canada

Address: Dept. of Math. and Stats.
Queen's University
Kingston, Ontario, Canada
K7L 3N6
Phone: 613 545 2395
Fax: 613 545 2964
Email: dmurdoch@mast.queensu.ca
Green, Peter, University of Bristol, Bristol, U.K.

Abstract: Propp and Wilson (1996) described a protocol, called coupling from the past*, for exact sampling from a target distribution using a coupled Markov chain Monte Carlo algorithm. In Murdoch and Green (1997) we described methods which can be used to apply coupling from the past to various MCMC samplers on continuous state spaces; rather than following the monotone sampling device of Propp and Wilson, our approach used methods related to gamma-coupling and rejection sampling to simulate the chain, and direct accounting of sample paths to check whether they had coalesced.*

In this presentation we will describe methods for coupling the random walk Metropolis-Hastings algorithm based on a pretty dissection of the area under the graph of a symmetric density function. These methods may be applied almost automatically to a wide class of Bayesian hierarchical models.

3. Exact Simulation of Markov Random Fields Using Coupling from the Past

Haggstrom, Olle, Chalmers University of Technology, Sweden

Address: Dept of Mathematics, Chalmers University of Technology, 412 96 Gothenburg, Sweden
Phone: +46 31 7725311
Fax: +46 31 161973
Email: olleh@math.chalmers.se
Nelander, Karin, Chalmers University of Technology

Abstract: *We discuss the possibility of using the Propp–Wilson coupling from the past (CFTP) approach to generate exact samples from various types of Markov random fields. The focus is mainly on examples rather than on general theory. One of our main messages is that the CFTP thechnique is applicable far beyond the class of stochastically monotone systems originally considered by Propp and Wilson. The talk is based on joint work with Karin Nelander.*

List of speakers who are nonmembers:

24.11 ims.11

IMS

Session Slot: 8:30-10:20 Thursday
Estimated Audience Size:
AudioVisual Request: Two Overheads

Session Title: Multiple Comparisons

Theme Session: No
Applied Session: No

Session Organizer: **Hsu, Jason**

Address:
Phone:
Fax:
Email:

Session Timing: 110 minutes total (Sorry about format):

Session Chair: **Spurrier, John** University of South Carolina

Address:
Phone:
Fax:
Email:

1. Semi-Bayesian Approaches to Multiple Pairwise Comparisons of Normal Means

Shaffer, Juliet P., University of California at Berkeley

Address: University of California, Department of Statistics, 367 Evans Hall # 3860, Berkeley, CA 94720-3860
Phone: 510-642-5411
Fax: 510-642-7892
Email: shaffer@stat.berkeley.edu

Abstract: Duncan proposed a Bayesian decision-theoretic approach to pairwise comparisons of normal means under the assumption that the population means are realizations of independent identically- distributed normal random variables; the method has minimum Bayes risk under the additional assumption of additive loss functions. Lewis and Shaffer have proposed pairwise comparison methods retaining the Duncan assumption of normally-distributed population means but using different loss functions. A modification of Lewis' method to control the familywise error rate (fwe), the most commonly-used error control criterion, allows a comparison with Shaffer's method, which controls the fwe for the overall null hypothesis only but is otherwise more compatible with the loss-function approach of Duncan. The comparison allows separate assessment of the increase in power due to (1) the assumption of a normal distribution of the population means over standard methods which do not incorporate this assumption, and (2) the modified criterion for error control.

2. Multiple Test Procedures for Safety Assessment

Tamhane, Ajit, Northwestern University

Address: Department of Statistics, 2006 Sheridan Road, Northwestern University, Evanston, IL 60208
Phone: (847)491-3577
Fax: (847)491-8005
Email: ajit@iems.nwu.edu

Abstract: We consider dose response studies for safety assessment of a compound and offer a new hypothesis testing approach for identifying the maximum dose level that is guaranteed to be safe with preassigned confidence. The conventional approach in which the lack of safety of a dose is demonstrated is also presented for comparison purposes. Both single-step (SS) and step-down (SD) multiple test procedures for identifying a maximum safe dose and a minimum unsafe dose are given in each case. These procedures control the familywise error rates for their respective hypotheses formulations. Their powers are studied via Monte Carlo simulation and sample sizes are provided for meeting specified power requirements. Procedures are illustrated by applying them to crop yield data from an agricultural field trial conducted to assess the safe level of a herbicide.

3. Stepwise Confidence Intervals without Multiplicity Adjustment

Hsu, Jason, The Ohio State University

Address:
Phone: 614-292-7663
Fax: 614-292-2096
Email: jch@stat.ohio-state.edu
Berger, Roger, North Carolina State University

Abstract: Not all simultaneous inferences need multiplicity adjustment. In this talk, we show that if the sequence of individual

inferences is predefined, and failure to achieve the desired inference at any step renders subsequent inferences unnecessary, then confidence sets for such inferences can be obtained without multiplicity adjustment. Dose response studies, designed to show superiority of treatments over a placebo (negative control) or a drug known to be efficacious (active control), fit into this setting. Toxicity studies, designed to show equivalence of treated groups with a placebo, also fit into this setting. It is shown that, for these studies, the confidence set approach is more reliable in generating simple methods with meaningful guarantee against incorrect decision than other approaches.

List of speakers who are nonmembers:

24.12 ims.12

IMS

Session Slot: 10:30-12:20 Thursday
Estimated Audience Size: 125-175
AudioVisual Request: Two Overheads

Session Title: Probability and Statistics in Telecommunications

Theme Session: No
Applied Session: No

Session Organizer: **Koshevnik, Yuly** MCI

Address: 2400 North Glenville Road, Richardson, TX 75082
Phone: (972) 918-6823
Fax: (972) 918-6257
Email: Yuly.Koshevnik@mci.com

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 5 minutes First Speaker - 20 minutes Second Speaker - 20 minutes Third Speaker - 20 minutes

Fourth Speaker - 20 minutes Discussant - 15 minutes (or none)
Floor Discusion - 5 minutes

Session Chair: **Koshevnik, Yuly** MCI

Address: 2400 North Glenville Road, Richardson, TX 75082
Phone: (972) 918-6823
Fax: (972) 918-6257
Email: Yuly.Koshevnik@mci.com

1. Moving Images: Semi- and Nonparametric Approach

Korostelev, Alexander, Wayne State University

Address: Department of Mathematics, F.A.B. 1247, Wayne State University, Detroit, MI 48202
Phone: (313) 577-3188
Fax: (313) 577-7596
Email: apk@math.wayne.edu

Abstract: On the basis of some examples, we discuss the peculiarities of the asymptotical large-sample analysis in nonparametric models of the moving images. As the asymptotical rates of convergence show, these problems cannot be treated as "snapshot" images in the "time-space" domains.

In the semiparametric image models, we want to estimate a finite dimensional parameter from noisy data in presence of an infinite dimensional "nuisance parameter" which describes the image shape. For example, if the evolution of the image is governed by some dynamical system, how accurately can the parameters of this system be restored? Unlike the classical semiparametric regression and density problems in which the root-n rates are typical under regularity conditions, in the image analysis the accuracy in estimating the finite dimensional parameter relates to the smooth functionals estimation. The smooth functionals of image such as the area, the center of gravity, etc., are known to have the intermediate rates of convergence comparing to those in parametric and nonparametric estimation. That's why we find the semiparametric image problems different from their parametric and nonparametric counterparts.

2. *Optimization of Polling Systems with Fast Service*

Kreinin, Alexander, Algorithmics, Inc.

Address: Algorithmics Inc. 822 Richmond Street West, Toronto, ON, Canada, M6J 1C9
Phone: (416) 703-0898
Fax: (416) 703-0767
Email: alex@algorithmics.com

Abstract: Polling systems with non-zero switchover time, intensive arrival flows, fast service and random discipline of service are studied in the talk. Under some assumptions, the process of queue lengths convergence to the trajectories of a dynamical system in the N-dimensional Euclidean space. We study the asymptotic behavior of the process, describe the properties of the limit dynamical system and obtain formulae for mean queue lengths expressed through the first moments of the input flow distributions, the distributions of service times and switchover times. We also derive the optimal random routing policy to minimize the average number of customers in the polling system.

3. *Monitoring of High Intensity Data Streams*

Yashchin, Emmanuel, IBM Research Center

Address: Thomas J. Watson IBM Research Center, P.O. Box 218, Yorktown Heights, NY 10598
Phone: (914) 945-1828
Fax: (914) 945-3434
Email: yashchi@watson.ibm.com

Abstract: This talk will discuss several problems related to analysis of data streams subject to abrupt changes in time (shifts, onset of drifts, etc.). Such data is frequently observed in a wide range of areas, such as quality control, communications and finance. It is typically described in terms of regimes and parameters that can be estimated, controlled or monitored. The talk will focus *on the problems of detection of unfavorable changes, estimation of the current level of parameters (filtering) and retrospective data analysis. It will present methodology based on the change-point theory for addressing these problems and give several examples of its application.*

4. *Interpretation Aids for Data-Rich Environments*

Hardy, William C., MCI Telecommunications Corp.

Address: 2400 North Glenville Road, Richardson, TX 75082
Phone: (972) 918-5925
Fax: (972) 918-6257
Email: Chris.Hardy@mci.com

Abstract: A data-rich environment is understood here to be one in which there is so much readily available data that it is possible to create samples with virtually no sampling variance. Such an environment is typified by the long-distance telecommunications industry, in which sample sizes of hundreds of thousands of links, millions of calls, billions of equipment operating hours, trillions of bits transmitted, etc. are the norm rather than the exception. This paper describes the shift in concerns from statistical to operational significance that occurs in such an environment, and illustrates analytical techniques that have been found to be useful in addressing those operational concerns with examples drawn from the author's experience as a telecommunications operations analyst.

Discussant: **Koshevnik, Yuly** MCI

Address: 2400 North Glenville Road, Richardson, TX 75082
Phone: (972) 918-6823
Fax: (972) 918-6257
Email: Yuly.Koshevnik@mci.com

List of speakers who are nonmembers: None

24.13 ims.13

IMS

Session Slot: 8:30-10:20 Tuesday
Estimated Audience Size:
AudioVisual Request: Two Overheads

Session Title: Probability and Finance

Theme Session: No
Applied Session: Yes

Session Organizer: **Kou, Steven** Univ. of Michigan and Columbia University

Address: Department of Statistics, Mason Hall, Ann Arbor, MI 48109-1027
Phone: 313-763-3498
Fax: 313-763-4676
Email: kou@umich.edu

Session Timing: 110 minutes total (Sorry about format):

Session Chair: **Kou, Steven** Univ. of Michigan and Columbia University

Address: Department of Statistics, Mason Hall, Ann Arbor, MI 48109-1027
Phone: 313-763-3498
Fax: 313-763-4676
Email: kou@umich.edu

1. Optimal Stopping and the Valuation of American Path-Dependent Options
Lai, T.L., Stanford University

Address: Department of Statistics, Stanford University, Stanford, CA 94305
Phone:
Fax:
Email: tlai@stat.stanford.edu

Abstract: *In this talk we first give a brief review of path-dependent options and present some recent results on corrected random-walk approximations to continuous-time optimal stopping problems involving Brownian motion. We then apply these results to evaluate the prices and early exercise regions of American path-dependent options.*

2. Probability Maximization and Active Portfolio Managment

Browne, Sid, Columbia University

Address: Graduate School of Business, Columbia University, New York, NY 10027
Phone:
Fax:
Email: sb30@columbia.edu

Abstract: *The objective of 'active' portfolio management is to beat a given index. Since an index is just some specific portfolio strategy, the 'active' portfolio manager is concerned with beating another (given) portfolio strategy. In this paper we consider the objective of maximizing the probability of beating another strategy. We thoroughly analyze the resulting probability maximizing strategy and compare its performance to more typical portfolio strategies used in practice. We also apply our results to other settings such as Risk Management, and the incorporation of external liabilities and income.*

3. Options Hedging and Statistical Uncertainty

Mykland, Per, University of Chicago

Address: Department of Statistics, University of Chicago, Chicago, IL 60637-1546
Phone:
Fax:
Email: mykland@galton.uchicago.edu

Abstract: What is the monetary value of statistical uncertainty? – There is a substantial body of theory on how to price and hedge options when the probabilistic model for the underlying security is known and continuous. There is also a big literature on statistical inference in such models. It is not clear, however, how statistical results can be used in pricing. In this talk, we explore two approaches to bridging this gap. On the one hand, we show that, to first approximation, one can offset the the statistical uncertainty by buying and selling the underlying security, thus putting an arbitrage based price on this uncertainty. On the other hand, prediction intervals can typically be converted to worst case hedging strategies. The latter device is particularly useful as it decreases the reliance on the probabilistic model. This reduces one's dependence on expert judgement in pricing this type of instrument, something which is greatly to be desired, both to improve management oversight and government regulation.

List of speakers who are nonmembers: none

24.14 ims.14

IMS

Session Slot: 2:00- 3:50 Tuesday
Estimated Audience Size: 125-175
AudioVisual Request: Two Overheads

Session Title: DATA MINING

Theme Session: No

Applied Session: No

Session Organizer: **Land, Stephanie** Carnegie Mellon University

Address: Department of Statistics Carnegie Mellon University Pittsburgh, PA 15213
Phone: (412) 268-1880
Fax: (412) 268-7828
Email: steph@stat.cmu.edu

Session Timing: 110 minutes total (Sorry about format):

> Opening Remarks by Chair - 2 minutes First Speaker - 30 minutes Floor Discusion - 5 minutes Second Speaker - 30 minutes Floor Discusion - 5 minutes Third Speaker - 30 minutes Floor Discusion - 8 minutes

Session Chair: **Land, Stephanie** Carnegie Mellon University

Address: Department of Statistics Carnegie Mellon University Pittsburgh, PA 15213
Phone: (412) 268-1880
Fax: (412) 268-7828
Email: steph@stat.cmu.edu

1. Probabilistic Models for Finding Patterns in Large Time-Series and Image Databases

Smyth, Padhraic, University of California, Irvine

Address: Information and Computer Science, University of California, Irvine, CA 92697-3425
Phone: (714) 824-2558
Fax: (714) 824 4056
Email: smyth@ics.uci.edu

Abstract: *Large databases containing multiple time series and multiple images are relatively commonplace across a variety of applications such as finance, medicine, remote-sensing, and*

engineering. In such applications it is common to want to match a prototype pattern of interest against the database, e.g., "find all images or time series segments in the database which are similar to this one". This is often referred to as "retrieval by content." Various different methods can be characterized by (1) representation (how the time series or image data are modeled), (2) similarity criteria (how similarity is quantified) and (3) search (how to avoid exponential time and space complexity when searching for patterns). Relatively little of this prior work is within a probabilistic framework, leading to ad hoc algorithms and systems. In this context, probabilistic models for time-series pattern matching are introduced. The probabilistic approach provides a natural semantics for describing prior knowledge on pattern deformations, and in turn implicitly defines a similarity metric. Results on a variety of databases demonstrate orders of magnitude speed-up over direct search.

2. Identifying Exceptions in Large Datasets

Ng, Raymond, University of British Columbia

Address: Department of Computer Science, 2366 Main Mall, Vancouver, B.C. V6T 1Z4, Canada
Phone: (604) 822-2394
Fax: (604) 822-5485
Email: rng@cs.ubc.ca

Abstract: "A person's noise is another person's signal." Most data mining techniques focus on finding patterns that fit the majority. Relatively little attention, except in the statistical community, has been given to the "dual" problem of finding exceptions.

In this talk, we shall consider two approaches for finding exceptions in large datasets with multiple attributes. The first approach is based on Tukey's definition of depth contours. We shall present algorithms for computing depth contours for 2 and 3 dimensions/attributes, and outline how to generalize the algorithms for even higher dimensions. The algorithms are designed to be able to deal with millions of objects/tuples residing on disks.

While the first approach can be considered to be order-based, the second approach is distance-based. Given a function defining the distance between every pair of objects, an object O is a (p,D)-outlier is there are at least p% of other objects in the dataset that are at least distance D away from O. We shall present algorithms for computing such outliers for high dimensional datasets. We shall compare and contrast the two approaches.

3. Very Sparse Structures for Caching Contingency-Tables Applied to Computationally Efficient Statistics and Datamining

Moore, Andrew, Carnegie Mellon University

Address: Robotics Institute EDSH 221 Carnegie Mellon University Pittsburgh, PA 15213
Phone: (412) 268-7599
Fax: (412) 268-5571
Email: awm@cs.cmu.edu

Abstract: In this talk we introduce new data structures and algorithms that can empirically give a 50-fold to 1000-fold computational speedup for machine learning methods such as Bayes net structure finders, rule learners and feature selectors operating on large datasets. Many machine learning algorithms need to do vast numbers of counting queries (e.g. "How many records have Color=Blue, Nationality=British, Smoker=False, Status=Married, Nose=Big?"). Similarly, many algorithms need to build and test many contingency tables. We provide methods for which, subject to certain assumptions, the costs of these kinds of operations can be shown to be independent of the number of records in the dataset and loglinear in the number of non-zero entries in the contingency table.

We provide a very sparse data structure, the ADTree (all-dimensions tree), to minimize memory use. We provide analytical worst-case memory and construction-cost bounds for this structure for several models of data distribution. We empirically demonstrate that tractably-sized data structures can be produced for large real-world datasets by (a)using a sparse tree structure

that never allocates memory for counts of zero, (b) never allocating memory for counts that can be deduced from other counts, and (c) not bothering to expand the tree fully near its leaves.

We show how the ADTree can be used to accelerate Bayes net structure finding algorithms, rule learning algorithms, and feature selection algorithms, and we provide a number of empirical results comparing ADTree methods against traditional direct counting approaches. We also discuss the possible uses of ADTrees in other machine learning methods, and discuss the merits of ADTrees in comparison with alternative representations such as kd-trees, R-trees and frequent sets.

List of speakers who are nonmembers:

24.15 ims.15

IMS

Session Slot: 8:30-10:20 Monday
Estimated Audience Size: 100
AudioVisual Request: Overhead Projector

Session Title: Difficult Likelihoods and Simple Solutions

Theme Session: No
Applied Session: No

Session Organizer: **Lele, Subhash** The Johns Hopkins University

Address: Department of Biostatistics, 615 N.Wolfe St., Baltimore, MD 21205.
Phone: (410) 955 3505
Fax: (410) 955 0958
Email: slele@welchlink.welch.jhu.edu

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 2 minutes First Speaker - 25 minutes Second Speaker - 25 minutes Third Speaker - 25 minutes Discussant - 10 minutes Floor Discusion - 5 minutes

Session Chair: **Lele, Subhash** The Johns Hopkins University

Address: Department of Biostatistics, 615 N. Wolfe St., Baltimore, MD 21205.
Phone: (410) 955 3505
Fax: (410) 955 0958
Email: slele@welchlink.welch.jhu.edu

1. A Composite Likelihood Approach to (Co)variance components estimation

Taper, Mark, The Montana State University

Address: Department of Biology, Montana State University, Bozeman, MT 59715.
Phone: (406) 994 2332
Fax:
Email: taper@rivers.oscs.montana.edu
Lele, Subhash, The Johns Hopkins University

Abstract: Variance and covariance components estimation has significant practical applications in animal breeding and evolutionary biology. The use of Maximum likelihood and Restricted maximum likelihood is difficult for very large data sets that commonly occur in animal breeding. This is mainly due to the necessity of inverting large matrices and the non-uniqueness of the solutions. In this paper we suggest a new approach that is applicable as generally as the maximum likelihood, involves no inversion of matrices and is guaranteed to have a unique solution. We present the consistency and efficiency properties of this method of estimation. We also present an analysis of a dataset that involves covariance components estimation in the presence of correlated random effects.

2. *Weighted likelihood equations: The case of mixture models*

Markatou, Marianthi, Columbia University

Address: Department of Statistics, 615 Mathematics Building, New York, NY 10027.
Phone: (212) 854 3969
Fax:
Email: markat@stat.columbia.edu

Abstract: Assume that the operating model is finite mixture of K components. We will discuss a computationally simple approach of obtaining estimates for the parameters of the mixture model which are consistent, efficient, asymptotically normal and robust in the presence of the outliers. We will discuss goodness of fit and address a number of issues associated with the algorithm such as starting values, stopping rule and rate of convergence of the algorithm.

3. *Pseudo-likelihood is interesting and easy to use*

Seymour, Lynne, University of Georgia

Address: Department of Statistics, 204 Statistics Building, Athens, GA 30602.
Phone: (706) 542 3307
Fax:
Email: seymour@rolf.stat.uga.edu

Abstract: In cases in which the joint distribution is intractable but conditional probabilities are known, the pseudo-likelihood provides a quick and dirty method for parameter estimation. The maximum psedo-likelihood parameter estimate has some very nice properties and is a promising alternative to the lengthy MCMC estimation schemes.

Discussant: **Natarajan, Nandini** Brown University

Address:

Phone:
Fax:
Email: ranjin@stat.brown.edu
List of speakers who are nonmembers: Taper, Mark

24.16 ims.16

IMS

Session Slot: 8:30-10:20 Thursday
Estimated Audience Size:
AudioVisual Request: Two Overheads

Session Title: Quasi Likelihood and Semiparametric Inference

Theme Session: No
Applied Session: No

Session Organizer: **Li, Bing** The Pennsylvania State University

Address: Department of Statistics The Pennsylvania State University 326 Joab Thomas Building University Park, PA 16802, USA
Phone: (814) 865 1952
Fax: (814) 863 7114
Email: bing@stat.psu.edu

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 5 or 0 minutes First Speaker - 30 minutes Second Speaker - 30 minutes Third Speaker - 30 minutes Discussant - Floor Discusion - 10 minutes (or 5 or 15)

Session Chair: **Shen, Xiaotong** Ohio State University

Address: Department of Statistics, Ohio State University
Phone:
Fax:

Email:

1. Projection, Hedging, and Pricing of Derivatives

McLeish, Donald, University of Waterloo

Address: Department of Statistics and Actuarial Science University of Waterloo Waterloo, Ontario, Canada N2L 3G1
Phone: (519) 888-4567, ext. 5534
Fax: (519) 664-1762
Email: dlmcleis@setosa.uwaterloo.ca
Kolkiewicz, Adam, University of Waterloo

Abstract: The use of projection techniques to hedge risk and price derivative securities is discussed. This permits either a semi-parametric or a non-parametric specification of models and approximate hedging, designed, for example, to limit the costs associated with continuously rebalancing a portfolio.

2. A Projected Quasi-Score Method to Reducing Sensitivity to Nuisance Parameters: Application to Measurement Error Models

Rathouz, Paul J., University of Chicago

Address: University of Chicago Dept. of Health Studies 5841 S. Maryland Ave., MC2007 Chicago, IL 60637
Phone: 773-834-1970 (o)
Fax: 773-701-1979
Email: rathouz@galton.uchicago.edu
Liang, Kung-Yee, Johns Hopkins University

Abstract: The projection method proposed by Waterman and Lindsay (1996) for desensitization to nuisance parameters was originally based on the likelihood function. An analagous method is developed for a quasi-likelihood regression set-up, simplifying model specification, assumptions, and the development of software for data analyses with many nuisance parameters in the mean function. The degree to which the full likelihood method extends to *the quasi-likelihood framework – and in what ways it falls short – is examined using finite sample properties of estimating functions. One application extends the work of Stefanski and Carroll (1987) to a broader class of measurement error models.*

3. On Quasilikelihood Equations with Nonparametric Weights

Li, Bing, The Pennsylvania State University

Address: Department of Statistics The Pennsylvania State University 326 Joab Thomas Building University Park, PA 16802, USA
Phone: (814) 865 1952
Fax: (814) 863 7114
Email: bing@stat.psu.edu

Abstract: To apply the quasi likelihood method one needs to know the mean and the variance function, so as to give more weight to the more accurate observations. If the variance function is unknown, then one should acquire the knowledge of the weights from the sample. One way to do so is by adaptive estimation, which in this context involves the consistent estimation of the entire variance function. However, for moderate sample sizes, such attempt may render the estimate too unstable, and we need a more flexible method which does not involve consistent estimation of the variance but still gives optimal weights to the extent to which a moderate sample size can tolerate. In this work we introduce a new type of quasi likelihood equation, whose weights belong to a smooth class that need not be large enough to be adaptive, but which are nevertheless optimal among the specified class. We will also present a rigorous development of the large sample properties of this equation, together with a sufficient and necessary condition for adaptivity.

List of speakers who are nonmembers: None

24.17 ims.17

IMS

Session Slot: 2:00- 3:50 Tuesday
Estimated Audience Size:
AudioVisual Request: Two Overheads

Session Title: Special Invited Papers

Theme Session: No
Applied Session: Yes

Session Organizer: **Lindsay, Bruce** Penn State University

Address: 422 Thomas Building, Department of Statistics, University Park, PA 16802
Phone: (814) 865-1220
Fax: (814) 863-7114
Email: bgl@psu.edu

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 5 First Speaker -45 minutes Second Speaker - 45 Floor Discusion - 10 minutes

Session Chair: **Li, Bing** The Pennsylvania State University

Address: Department of Statistics The Pennsylvania State University 326 Joab Thomas Building University Park, PA 16802, USA
Phone: (814) 865 1952
Fax: (814) 863 7114
Email: bing@stat.psu.edu

1. Shapes, Directions and Images
Mardia, Kanti V., University of Leeds

Address: Department of Statistics University of Leeds Leeds, LS2 9JT, England
Phone: 0113 233 5100
Fax: 0113 233 5102
Email: K.V.Mardia@leeds.ac.uk

Abstract: *In general, the methods of multivariate analysis break down for data in non-Euclidean spaces. Directions and shapes are two prime examples. A shape is described by the information in a set of landmarks which is invariant under the transformations of location, rotation and scaling, and shape analysis concentrates on methods which respect these invariances and preserve geometrical properties of object. We deal with statistical methodolgy to measure, describe and compare the shapes of objects; of particular importance is the need to visualize shape variability. Considerations are similar to the well established topic of directional data analysis but somewhat more complex.*

This talk coincides with the 80th anniversary of the famous classic book "Growth and Form" of D'Arcy Thompson which raised basic biological questions about shape analysis. We describe the latest developments in shape space, Procrustes methods, tangent approximations, shape distributions, symmetry in shape, image warping and averaging, and object recognition. Practical examples will be given from various fields including medical imaging, face analysis and biology.

2. Statistical Analysis of Shape for Pattern Differentiation and Recognition

Rao, C. R., Pennsylvania State University

Address: 326 Joab Thomas Building, University Park, PA 16802
Phone: 814-865-3194
Fax: 814-863-7114
Email: crr1@psu.edu

Abstract: *The shape of an object, data set or image can be defined as the total of all information that is invariant under translations, rotations, and isotropic rescaling. Thus two objects are said to have the same shape if one object can be exactly fitted into the other by the above transformations.*

The last decade has seen a considerable growth in interest in the statistical theory of shape. There are different groups of research workers headed by Bookstein, Kendall and Mardia who

advocate different methods of shape analysis. An attempt will be made to prove a unified treatment by defining shape measurements which are independent of scale, translation and rotation, and developing methods of statistical inference based on them. If an object is characterized by certain landmarks, then Euclidean distances between landmarks as suggested by Lele, or the angles of triangles obtained by suitable triangulation of landmarks consitute shape measurements on which parametric or nonparametric statistical methods could be applied. If landmarks are not available, geometrical properties like curvature of the boundary of objects provide shape measurements for which appropriate statistical methods could be developed.

A number of examples with live data will be presented illustrating the use of shape measurements in pattern recognition and tests of hypotheses on shape differences.

List of speakers who are nonmembers: None

24.18 ims.18

IMS

Session Slot: 8:30-10:20 Wednesday
Estimated Audience Size: 125-175
AudioVisual Request: None

Session Title: Statistics and Genetics

Theme Session: No
Applied Session: Yes

Session Organizer: **McPeek, Mary Sara** Department of Statistics, University of Chicago

Address: Department of Statistics, 5734 S. University Ave., Chicago, IL 60637
Phone: 773-702-7554
Fax: 773-702-9810

Email: mcpeek@galton.uchicago.edu

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 0 minutes First Speaker - 35 minutes Second Speaker - 35 minutes Third Speaker - 35 minutes Discussant - none Floor Discusion - 5 minutes

Session Chair: **McPeek, Mary Sara** Department of Statistics, University of Chicago

Address: Department of Statistics, 5734 S. University Ave., Chicago, IL 60637
Phone: 773-702-7554
Fax: 773-702-9810
Email: mcpeek@galton.uchicago.edu

1. Mapping disease genes using distantly related individuals

Feingold, Eleanor, Department of Human Genetics, University of Pittsburgh

Address: Department of Human Genetics, 130 DeSoto Street, Crabtree Hall A310, Pittsburgh, PA 15260
Phone: 412-383-8599
Fax: 412-624-3020
Email: efein@helix.hgen.pitt.edu
Durham, L. Kathryn, Imperial Cancer Research Fund, U.K.

Abstract: One standard method for mapping human disease genes is to collect families in which two or more members are affected with the disease, and search for regions where the affected relatives share genetic material (identity by descent). More recently, success has been seen with "linkage disequlibrium" mapping, which uses putatively unrelated individuals from a population. Both of these methods have been studied from a statistical point of view, in differing degrees of depth. We consider the statistical problems that arise in an intermediate situation: a study that searches for shared genome segments among distant relatives, whose exact relationship may be known or unknown.

2. Bayesian haplotype analysis for linkage disequilibrium

Liu, Jun, Department of Statistics, Stanford University

Address: Department of Statistics, Stanford University, Stanford, CA 94305
Phone: 650-723-2623
Fax: 650-725-8977
Email: jliu@stat.stanford.edu
Teng, Jun, Stanford University
Risch, Neil, Stanford University

Abstract: *Haplotype analysis of disease chromosomes can help identify probable historical recombination events and consequently localize the disease gene. Available analyses only use marginal and pairwise allele frequency information. In this talk we propose a Bayesian framework that helps in utilizing full haplotype information overcoming various complications such as missing data, multiple ancestral clusters, and possible data contaminations.*

3. Assessing linkage disequilibrium by the decay-of-haplotype-sharing method

McPeek, Mary Sara, Department of Statistics, University of Chicago

Address: Department of Statistics, 5734 S. University Ave., Chicago, IL 60637
Phone: 773-702-7554
Fax: 773-702-9810
Email: mcpeek@galton.uchicago.edu
Strahs, Andrew, University of Chicago
DiRienzo, Anna, University of Chicago
Hall, Diana, University of Chicago

Abstract: *The extent of linkage disequilibrium in a genomic region for a given population reflects such factors as isolation of*

the population, historical changes in population size, and natural selection at the locus. Populations which tend to show a high degree of linkage disequilibrium are expected to be particularly useful for genetic mapping. Most previous measures of linkage disequilibrium have been based on pairwise locus data. We have developed a method for assessing linkage disequilibrium based on analysis of multilocus haplotypes. The method models the rate of decrease of haplotype sharing with increasing distance. In addition to measuring linkage disequilibrium, it also has the potential to detect genes by population association. We have applied the method to compare the extent of linkage disequilibrium in two genomic regions across several populations, including Italians, Ashkenazim, Hutterites, African Americans, Gambians, and Bamileke. The results are directly applicable to development of strategies for mapping complex traits in various human populations.

List of speakers who are nonmembers: None

24.19 ims.19

IMS

Session Slot: 2:00- 3:50 Wednesday
Estimated Audience Size:
AudioVisual Request:

Session Title: Special Invited Papers II

Theme Session: No
Applied Session: No

Session Organizer: **Lindsay, Bruce** The Pennsylvania State University

Address: 422 Thomas Building, Department of Statistics, University Park, PA 16802
Phone: (814) 865-1220
Fax: (814) 863-7114
Email: bgl@psu.edu

Session Timing: 110 minutes total (Sorry about format):

Session Chair: **Lindsay, Bruce** The Pennsylvania State University

Address: 422 Thomas Building, Department of Statistics, University Park, PA 16802
Phone: (814) 865-1220
Fax: (814) 863-7114
Email: bgl@psu.edu

1. Inference in High Dimension Models

Murphy, Susan, University of Michigan

Address: The University of Michigan 1440 Mason Hall Ann Arbor, MI 48109-1027
Phone: (313) 763-3519
Fax: (313) 763-4676
Email: murphy@stat.psu.edu
van der Vaart, A.W., Free University, Amsterdam

Abstract: Likelihood inference in high dimensional models appears, at best, to be characterized by variety: variety in the form of the likelihood, variety in distributional properties, variety in rates of convergence, variety in the norms, etc. Even in apparently "regular/smooth" models, there does not appear to be any generally applicable approaches for ascertaining asymptotic properties. This state of affairs is reminiscent of maximum likelihood estimation in smooth parametric models during the pre-Cramér days.

This talk will discuss examples illustrating the above variety, pointing out areas in much need of research. In models in which a finite dimensional parameter is of primary interest, it is possible to give a general methodology for ascertaining asymptotic properties of inferences based on a maximum likelihood estimator. This methodology assumes smoothness similar to Cramér's

classical conditions and can be applied to a wide variety of models which are not smooth in the nuisance parameter. Aspects which need further development will be discussed.

List of speakers who are nonmembers:

24.20 ims.20

IMS

Session Title: High-dimensional numerical integration

Session Slot: 10:30-12:20 Thursday
Estimated Audience Size: 30-50
AudioVisual Request: None or Two Overheads
Theme Session: No
Applied Session: No

Session Organizer: **Stein, Michael** University of Chicago

Address: Department of Statistics, 5734 University Ave., Chicago, IL 60637
Phone: 773-702-8326
Fax: 773-702-9810
Email: stein@galton.uchicago.edu

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 0 minutes First Speaker - 30 minutes Second Speaker - 30 minutes Third Speaker - 30 minutes Discussant - 10 minutes Floor Discusion - 10 minutes

Session Chair: **Stein, Michael** University of Chicago

Address: Department of Statistics University of Chicago 5734 University Ave. Chicago, IL 60637
Phone: 773-702-8326

Fax: 773-702-9810
Email: stein@galton.uchicago.edu

1. *Latin supercube sampling*

Owen, Art, Stanford University

Address: Department of Statistics Stanford University Stanford, CA 94305
Phone:
Fax:
Email: art@stat.Stanford.EDU

Abstract: Some very high dimensional integrands arise in problems such as graphical rendering, particle transport, computational finance and queuing systems. The methods of choice for such problems are usually Monte Carlo (MC) or quasi-Monte Carlo (QMC). In theory QMC methods are far more accurate than MC, at least asymptotically, but they can lose their effectiveness in high dimensions. Recent work has centered on using QMC on a small set of key dimensions and MC on the others. This talk considers how to use QMC on more than one set of input dimensions, by randomizing the run order within each set, generalizing the method of Latin hypercube sampling. In combination with methods of engineering the integrand to have more low dimensional structure, the result is to extend the reach of QMC methods to harder problems, including some with infinite dimension. When the QMC methods themselves are randomized it is possible to get a data based estimate of the accuracy of the integration.

2. *Adaptive Integration and Analysis of Average Errors*

Ritter, Klaus, University of Erlangen-Nurnberg

Address: Mathematisches Institut Universität Erlangen-Nürnberg Bismarckstraße 1 1/2 91054 Erlangen Germany
Phone:
Fax:
Email: ritter@@mi.uni-erlangen.de

Müller-Gronbach, Thomas, Freie Universität Berlin

Abstract: *We discuss numerical integration of functions with inhomogeneous and a priori unknown local smoothness. For such problems adaptive (sequential) methods are a natural choice. One tries to detect regions of different local smoothness and adjusts further sampling accordingly.*

In a worst case analysis, no method can successfully deal with unknown local smoothness. This result holds true under very mild assumptions.

Hence we study average errors with respect to probability measures on function spaces. We present new adaptive algorithms and derive optimality properties in the average case setting. Our analysis proves superiority of adaptive methods and it comes close to some heuristics that are used in the literature.

3. (T,M,S)-Nets and Their Applications in Numerical Integration

Mullen, Gary, Pennsylvania State University

Address: Department of Mathematics The Pennsylvania State University University Park, PA 16802

Phone:

Fax:

Email: mullen@math.psu.edu

Abstract: *For a variety of applied problems one needs to be able to distribute points as uniformly as possible in the unit cube* $I = [0,1)^s$*. There are various ways to try and do this but one of the most effective is through the use of* (t,m,s)*-nets in base* b*. Such a net is a collection of* b^m *points in* I *so that each "elementary interval" of volume* $1/(b^{m-t})$ *contains its fair share of points, namely* b^t*.*

In this talk we will briefly discuss some of the known constructions for (t,m,s)*-nets. As an illustration in numerical integration, we will provide a comparison of the effectivenss of several*

methods of choosing node points including the use of nets, Monte Carlo methods, and Sobol sequences.

Discussant: **Stein, Michael**

Address:
Phone:
Fax:
Email:
List of speakers who are nonmembers: Ritter, Mullen

24.21 ims.21

IMS

Session Slot: 4:00- 5:50 Sunday
Estimated Audience Size:
AudioVisual Request: Two Overheads

Session Title: Local Learning: Modern Nonparametric Regression and Classification

Theme Session: No
Applied Session: No

Session Organizer: **Tibshirani, Rob** University of Toronto, Toronto, Canada

Address: Dept of Preventive Med & Biostats, and Dept of Statistics Univ of Toronto, Toronto, Canada M5S 1A8.
Phone: 416-978-4642 (PMB), 416-978-0673 (stats)
Fax: 416 978-8299
Email: tibs@utstat.toronto.edu

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 5 minutes First Speaker - 25 minutes Second Speaker - 25 minutes Third Speaker - 25 minutes Discussant - 10 minutes Floor Discussion - 10 minutes

Session Chair: **TBN**

Address:
Phone:
Fax:
Email:

1. ***Theory and practice of boosting***
Schapire, Robert, AT&T Labs

Address: 180 Park Avenue, Room A279 Florham Park, NJ 07932-0971
Phone: 973-360-8329
Fax: 973-360-8970
Email: schapire@reseach.att.com

Abstract: Boosting is a general method for producing a very accurate classification rule by combining rough and moderately inaccurate "rules-of-thumb." Any given classification method can be used to find these weak rules-of-thumb. Boosting works by running the given classification method many times, each time on a different set of training examples, and then combining the resulting rules-of-thumb.

In this talk, I will introduce the most recent boosting algorithm, called AdaBoost, explain the basic, underlying theory of boosting, and review some experiments comparing AdaBoost to other algorithms, including Breiman's "bagging" algorithm. Besides demonstrating that AdaBoost often performs well in practice, several of those who experimented with AdaBoost (including Drucker and Cortes, Quinlan, and Breiman) made the surprising observation that the algorithm usually does not suffer from overfitting, even when it generates models that are very complex. I will describe a new theoretical analysis which explains this phenomenon.

[This talk includes joint work with Yoav Freund, Peter Bartlett and Wee Sun Lee.]

2. ***Locally Bagged Decision Trees***
Rao, J.S., Cleveland Clinic

Address: Department of Biostatistics The Cleveland Clinic 9500 Euclid Ave, Cleveland, OH 44195
Phone: 216-445-7844
Fax: 216-444-8023
Email: srao@bio.ri.ccf.org
Potts, W.J.E., SAS Institute Inc.

Address: SAS Institute Inc. SAS Campus Drive, Bldg H Cary, NC 27513
Phone: 919-677-8000 ext. 4629
Fax: 919-677-8225
Email: saswzp@wnt.sas.com

Abstract: *Tree-structured classifiers (CART, C4.5) are attractive in that they produce flexible models of simple structure. They are however unstable in that small perturbations of the training data can give different trees. The idea of bootstrap aggregation of trees (bagging) was introduced by Breiman and can lead to greatly improved predictions but loss of interpretable structure. Under the hypothesis that instability may be local in nature, we introduce a version of bagging that uses bootstrap resampling at a node along with a determination of local instability using principal co-ordinate analysis. The result is a new class of models with bagging-like improvements in prediction accuracy, and that maintain a tree-like topology. The method is illustrated on benchmark datasets from the University of California-Irvine machine learning repository.*

3. Minimum Description Length and Extended Linear Modeling

Hansen, Mark, Bell Labs, Lucent Technologies

Address:
Phone:
Fax:
Email:

Abstract: *In this talk, we investigate the use of the principle of minimum description length (MDL) for problems of model*

selection in the context of an extended linear model. By viewing statistical modeling as a means of generating descriptions *of observed data, the MDL framework discriminates between competing models based on the* complexity *of each description. In our application, we will consider a variety of spline bases (including classical polynomial and smoothing splines) as building blocks and employ MDL to identify promising models.*

This is joint work with Bin Yu, University of California at Berkeley.

List of speakers who are nonmembers:

24.22 ims.22

IMS

Session Slot: August 12 8:30- 10:20 Tuesday
Estimated Audience Size: 125-175
AudioVisual Request: Two Overheads

Session Title: Conditional Independence Properties of Structural Equation Models

Theme Session: No
Applied Session: No

Session Organizer: **Tritchler, David** Ontario Cancer Institute and University of Toronto

Address: Ontario Cancer Institute 610 University Avenue Toronto, Ontario M5G 2M9 Canada
Phone: (416) 946 2064
Fax: (416) 946 2024
Email: tritchle@oci.utoronto.ca

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 0 minutes First Speaker - 30 minutes Second Speaker - 30 minutes Third Speaker - 30 minutes Discussant - 15 minutes Floor Discusion - 5 minutes

Session Chair: **Tritchler, David** Ontario Cancer Institute and University of Toronto

Address: Ontario Cancer Institute 610 University Avenue Toronto, Ontario M5G 2M9 Canada
Phone: (416) 946 2064
Fax: (416) 946 2024
Email: tritchle@oci.utoronto.ca

1. Marginalization and Conditioning in Graphical Models
Koster, Jan, Erasmus University Rotterdam

Address: Erasmus University Rotterdam Department of Sociology P.O. Box 1738 Rotterdam NL-3000 DR The Netherlands
Phone: +31.10.408.20.77
Fax: +31.10.452.58.70
Email: koster@soc.fsw.eur.nl

Abstract: *A class of graphs is introduced which is closed under marginalization and conditioning. It contains all graphs associated with linear structural equation models. Using earlier consistency results, conditional independence properties of subsystems of linear structural equation models are discussed.*

2. Analyzing Structural Underdetermination and Model Equivalence Via Graphical Models
Richardson, Thomas, University of Washington

Address: University of Washington Department of Statistics P.O. Box 35-4322 Seattle, WA 98195-7232
Phone: +1 206 685 8488
Fax: +1 206 685 7419
Email: tsr@stat.washington.edu
Spirtes, Peter, Carnegie-Mellon University

Abstract: *Analyses based on structural equation models often claim to provide insight into the nature of the mechanism which generated a given set of data. A fundamental criticism of structural equation modelling is that, without background knowledge, there may be many plausible data generating structures which give rise to statistically equivalent models. For example, there may be two SEM models which fit the data well, in one of which* $A \to B$, *while in the other* $B \to A$. *These problems are particularly severe in cases where there may be unmeasured confounding variables. One approach to addressing this problem is to consider equivalence classes of structural equation models. This talk will describe recent work which uses graphical methods to characterize equivalence classes of structural equation models with latent variables and/or correlated errors.*

3. Testing the Testable and Identifying the Identifiable in Structural Equation Models

Pearl, Judea, UCLA

Address: UCLA Department of Computer Science Room 4515 BH Los Angeles CA 90024
Phone: (310) 825 3243
Fax: (818) 789 6311
Email: judea@cs.ucla.edu

Abstract: *Recent progress in graphical models has brought clarity, precision and legitimacy to structural equation modeling (SEM). Fundamental concepts that were left ambiguous by decades of inadequate formulations can now be given formal and operational meaning . What makes a set of equations "structural," what assumptions should be ascertained by the authors of such equations, what policy claims are advertised by a given set of structural equations, and what portions of these claims are testable are some of the questions that now receive simple and mathematically precise answers. The talk will survey graphical techniques for identifying the testable claims of SEM, for recognizing the identifiable parameters in a given SEM, and for de-*

termining when a variable is exogenous, when a variable is an "instrument," and what those "so-called disturbance terms" are.

Discussant: **TBN**

Address:
Phone:
Fax:
Email:
List of speakers who are nonmembers: None

24.23 ims.23

IMS

Session Slot: 4:00- 5:50 Sunday
Estimated Audience Size: 150
AudioVisual Request: None

Session Title: Observational Studies

Theme Session: No
Applied Session: No

Session Organizer: **Wasserman, Larry** Carnegie Mellon

Address: Dept. of Stat, Carnegie Mellon
Phone: 412-268-8727
Fax: 412-268-7828
Email: larry@stat.cmu.edu

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 0 minutes First Speaker - 30 minutes Second Speaker - 30 minutes Third Speaker - 30 minutes Discussant - 0 minutes Floor Discusion - 15 minutes

Session Chair: **Wasserman, Larry** Carnegie Mellon

Address: Dept. of Stat, Carnegie Mellon
Phone: 412-268-8727
Fax: 412-268-7828
Email: larry@stat.cmu.edu

1. Sensitivity Analysis in Observational Studies: Two Simple Examples

Rosenbaum, Paul, Univ. of Pennsylvania

Address: Paul R. Rosenbaum
The Wharton School
University Of Pennsylvania
Philadelphia, PA 19104-6302
Phone: 215-898-3120
Fax: 215-898-1280
Email: rosenbap@wharton.upenn.edu

Abstract: In an observational study, treatments are not randomly assigned to experimental subjects, so treated and control subjects may differ prior to treatment in ways that matter for the outcomes under study. Even when treated and control subjects are matched on recorded covariates, they may nonetheless differ in ways that have not been recorded. A sensitivity analysis asks how hidden biases of various magnitudes might alter the conclusions of a study. Two simple examples of sensitivity analysis will be presented, one from epidemiology and one from economics. In the economics example, the sensitivity of an instrumental variable estimator will also be examined.

2. Experiments, Randomized Trials, Observational Studies, and Causal Inference

Scheines, Richard, Carnegie Mellon University

Address: Dept. of Philosophy, CMU, Pittsburgh, PA 15213
Phone: 412-268-8571
Fax: 412-268-1440

Email: R.Scheines@andrew.cmu.edu

Abstract: Causal Inference from controlled experiments and large randomized clinical trials is relatively uncontroversial, whereas inference from uncontrolled observational studies is almost taboo. In this paper I use graphical causal models to show how the structure of inference is similar in all of these contexts. I show the assumptions needed to infer that a relationship is causal, how controlled experiments and randomized trials give us reason to endorse these assumptions, and how patterns of independence can do the same job in observational studies.

3. A New Approach to Observational Studies of Disease-specific Mortality

Joffe, Marshall, University of Pennsylvania

Address: Department of Biostatistics and Epidemiology
602 Blockley Hall
423 Guardian Drive
University of Pennsylvania School of Medicine
Philadelphia, PA 19104-6021
Phone: 215- 573-7395
Fax: 215- 573-4865
Email: joffe@cceb.med.upenn.edu

Abstract: In observational research, investigation of possible causal links between a treatment or exposure and outcome often rely on subjects' self-report regarding exposure or treatment received; obtaining this information can be expensive. When the outcome event is rare, it is more efficient to obtain such self-report from all subjects who experience the event and a randomly chosen subset of the rest of the study cohort. Such case-control designs are problematic for mortality outcomes, because dead subjects cannot report information on previous exposures. For cause-specific mortality, a feasible hybrid design (Weiss and Lazovich, 1996) can often preserve much of the cost efficiency of case-control design. In this design, exposure is assessed at the time of disease

diagnosis for all subjects diagnosed with the disease of interest, and for a subset of the remaining subjects. Unfortunately, analyzing data gathered this way as if it were standard case-control data can give biased results. When, as is usually the case with cancer outcomes, only a small proportion of potential subjects will be diagnosed with the condition of interest, an alternate method of analysis of such data is possible. This approach, which is based on formal definitions of the effect of a time-varying treatment and on the G-computation algorithm (Robins, 1986), may be applied even when the sampling fraction of non-diseased subjects is not known. We illustrate and compare conventional and new methods with data from a hypothetical study of breast cancer mortality.

List of speakers who are nonmembers: None

24.24 ims.24

IMS

Session Slot: 2:00- 3:50 Monday
Estimated Audience Size: 125-175
AudioVisual Request: None or Two Overheads

Session Title: Information Theory

Theme Session: No
Applied Session: No

Session Organizer: **Yu, Bin** Univ of California at Berkeley and Bell Labs, Lucent Technologies

Address: Statistics Department, 367 Evans Hall, #3860, University of California, Berkeley, CA 94720-3860
Phone: (510) 642-2021
Fax: (510) 642-7892
Email: binyu@stat.berkeley.edu

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 5 or 0 minutes First Speaker - 30 minutes (or 25) Second Speaker - 30 minutes Third Speaker - 30 minutes Discussant - 10 minutes (or none) Floor Discusion - 10 minutes (or 5 or 15)

Session Chair: **Hansen, Mark** Bell Labs, Lucent Technologies

Address:
Phone:
Fax:
Email: cocteau@research.bell-labs.com

1. Information Theoretic Methods in Statistics

Csiszár, Imre, Mathematics Institue, Hungary Academy of Science

Address: Math. Inst. Hungar. Acad. Sci H1364 Budapest P.O.Box 127, Hungary
Phone: 011 (36-1) 117-7175
Fax: 011 (36-1) 177-7166
Email: csi@math-inst.hu

Abstract: *While Information Theory uses to a large extent the concepts and methods of Probability and Statistics, also the other way round, concepts and methods developed in Information Theory have been fruitfully applied to various problems of Probability and Statis- tics. In this talk some such applications will be reviewed, ranging from Hajek's (1957) information theoretic proof of the dichotomy theorem for Gaussian measures to Marton's (1997) results on measure concentration via her information-theoretic method. Topics covered will include large deviations theory, hypothesis testing, non-para- metric estimation. The two major inference principles motivated by Information Theory, viz. Maximum Entropy /or Minimum Discrimination Information/ and Minimum Description Length would require more time to discuss at any depth. These will be but briefly illustrated via relevant examples.*

2. Model Selection and Hypothesis Testing by the MDL Principle

Rissanen, Jorma, IBM Research Division

Address: IBM Research Division, ARC DPE-B2/802, San Jose, CA 95120-6099
Phone: (408) 927-1813
Fax: (408) 927-2100
Email: rissanen@almaden.ibm.com

Abstract: The central idea of the MDL (Minimum Description Length) principle, both in model selection and hypothesis testing, is to represent a class of models (hypotheses) by a a universal model capable of imitating the behavior of any model in the class. The principle then calls for a model class whose representative assigns the largest probability or density to the observed data. Two examples of universal models for parametric classes $\mathcal{M}$ are the normalized maximum likelihood function

$$\hat{f}(x^n|\mathcal{M}) = f(x^n|\hat{\theta}(x^n))/\int_{\Omega} f(y^n|\hat{\theta}(y^n))dy^n,$$

where Ω is a small but simple-to-describe set making the integral finite, and a mixture

$$f_w(x^n|\mathcal{M}) = \int f(x^n|\theta)w(\theta)d\theta$$

as a convex linear functional of the models. A Bayes factor $B = f_w(x^n|\mathcal{M}_1)/f_v(x^n|\mathcal{M}_2)$ in this interpretation is the ratio of mixture representatives of two model classes, and the Bayesian model comparison reduces to a particular case of the MDL principle. However, mixtures need not be the best representatives.

In hypothesis testing, such as in a test of a gaussian singelton class $H_0 = \{f(x^n|\theta_0, \sigma)\}$ against the remaining class $H_1 = \{f(x^n|\theta, \sigma)\}$, we may look for a representative of H_1 such that the ratio of its power to the worst case power of the members in H_1 differs least from unity. The goodness of how well the two hypotheses are separated may then be assessed by the power

of such a minmax representative. We show that the normalized maximum likelihood is a minmax representative, while no mixture, taken with respect to $w(\theta)$ which is symmetric about θ_0 and nonincreasing in $|\theta - \theta_0|$ exists with the same minmax property.

Time permitting we also discuss other applications of the MDL principle to model selection.

3. Information and the Clone Mapping of Chromosomes

Yu, Bin, Univ of California at Berkeley and Bell Labs, Lucent Technologies

Address: Statistics Department, 367 Evans Hall, #3860, University of California, Berkeley, CA 94720-3860
Phone: 510-642-2021
Fax: 510-642-7892
Email: binyu@stat.berkeley.edu
Speed, Terry, Univ of California at Berkeley

Abstract: The first step in the Human Genome Project is to assemble DNA fragments, called clones, to form clone maps, which allow the detailed study of chromosomal regions of biological interest. In this talk, we answer biologist Lehrach's question about how much information is needed to complete a clone map by formulating a number of different notions of a clone map. The entropy of each notion (or configuration variable) is tightly bounded. Our results are useful for planning future mapping efforts. In particular, it follows that the cosmid clone mapping for the roundworm requires about 40 times as much information as that for the bacterium E. coli*, and that such mapping for humans requires about 1,500 times as much information as that for the bacterium* E. coli.

List of speakers who are nonmembers:

24.25 ims.25

IMS

Session Slot: 8:00- 10:00 Monday
Estimated Audience Size: xxx
AudioVisual Request: Two Overheads

Session Title: IMS President's Invited Address

Theme Session: No
Applied Session: No

Session Organizer: **TBN**

Address:
Phone:
Fax:
Email:

Session Timing: 110 minutes total (Sorry about format):

usual

Session Chair: **TBN**

Address:
Phone:
Fax:
Email:

1. Mysticism in Statistics: From the Bible Codes to the Cult of Fisher

Diaconis, Persi, Cornell University

Address: Department of Mathematics, ORIE, Cornell University
Phone:
Fax:
Email: persi@orie.cornell.edu

Abstract: *Statistics is a different subject than mathematics or probability. Seeking for its essence leads some to follow gurus like Fisher or Definetti. The Bible Codes is a recent example of how believers can go off in crazy directions; I will review this and other examples.*

List of speakers who are nonmembers:

Chapter 25

Statistical Society of Canada (2)

Program Chair: **Binder, David**

25.1 ssc.01

Sponsoring Society: SSC; Co-sponsored by ASA Section on Survey Research Methods

Session Slot: 2:00- 3:50 Wednesday
Estimated Audience Size: ???
AudioVisual Request: None

Session Title: Innovative Methodology in Government Surveys

Theme Session: Yes
Applied Session: Yes

Session Organizer: **Brackstone, Gordon J.** Statistics Canada

Address: Statistics Canada 26-J R.H. Coats Building Ottawa, Ontario K1A 0T6 CANADA
Phone: 613-951-9908
Fax: 613-951-4842

Email: bracgor@statcan.ca

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 5 minutes First Speaker - 30 minutes Second Speaker - 30 minutes Third Speaker - 30 minutes Discussant - 10 minutes (or none) Floor Discussion - 5 minutes

Session Chair: **Binder, David** Statistics Canada

Address: 11-A RH Coats Bldg. Business Survey Methods Division Statistics Canada Ottawa, ON K1A OT6 CANADA
Phone: (613) 951-0980
Fax: (613) 951-1462
Email: binddav@statcan.ca

1. The Workplace and Employee Survey: A New approach to Linking Employer and Employee Data

Patak, Zdenek, Statistics Canada

Address: Statistics Canada 11-J R.H. Coats Building Ottawa, Ontario K1A 0T6 CANADA
Phone: 613-951-9434
Fax: 613-951-1462
Email: patazde@statcan.ca
Hidiroglou, Michael, Statistics Canada
Morel, José, Statistics Canada
Lavallée, Pierre, Statistics Canada

Abstract: The Workplace and Employee Survey (WES) is designed to be a longitudinal survey. Its objective is to investigate relationships among competitiveness, innovation, technology use and human resource management on the employer side and technology use, training, job stability and earnings on the employee side. The first wave sample uses a two-stage design for the employer portion, and a third stage for the employee portion. Various sampling options were considered for subsequent waves. Variance components for each stage of selection are computed by *extending the functionality of the Generalized Estimation System (GES). This system was originally built to handle single stage designs. WES uses the Generalized Sampling System (GSAM) for sample selection at each wave. The use of these systems simplifies the survey process and improves its timeliness.*

2. Recent Developments in the American Community Survey

Alexander, Charles H., Bureau of the Census

Address: Bureau of the Census Demographic Statistical Methods Division Room 3705-3 Washington, D.C. 20233
Phone: 301-457-4290
Fax: 301-457-3851
Email: ???

Abstract: The American Community Survey (ACS) will be a nationwide rolling sample survey that will provide intercensal social and economic data for "communities" (places and population groups) of all sizes starting after Census 2000, using multi-year moving averages for the smallest communities. It will replace the "long form" content sample in the 2010 and future censuses. The ACS is being tested in four sites that started in 1996, with additional sites added in 1997 and 1998.

This paper presents the latest results about the quality and potential uses of the ACS data, based on analyses of results from the test sites. This includes new details on issues related to the use of multi-year averages, that were confronted in preparing to release the combined 1996-1997 estimates.

3. How Can a Government Agency Measure Child Development: The Experience of the Canadian National Longitudinal Survey of Children and Youth

Michaud, Sylvie, Statistics Canada

Address: Social Survey Methods Division, 7-C2, JT Bldg, Ottawa, Ontario, K1A 0T6
Phone: 613-951-9482

Fax: 613-951-3253
Email: michsyl@statcan.ca

Abstract: *The Canadian National Longitudinal Survey of Children and Youth (NLSCY) is an ambitious survey undertaken jointly by Human Resources Development Canada and Statistics Canada. In 1994, a sample of children aged between 0-11 years was selected and the cohort has since been interviewed every second year. Child development is measured through the collection of information from various sources. Detailed information about the child and the family environment is obtained through the person most knowledgeable in the household. When the child is four or five years old, a vocabulary test is administered in the home interview. Starting at age ten, the child/adolescent is given a self-administered questionnaire that is completed while the most knowledgeable person is being interviewed. There are some tests administered in the school (starting roughly at age 7). More details about the school environment and the children's behavior at school is obtained by a questionnaire filled in by the teacher and the principal. The development of such a holistic approach to child development requires trade-offs in terms of content. The content varies through time; some measures (such as scales on behavior or temperament) can be only partially administered due to the eventual length of the interview, and choices have to be made as to what is repeated longitudinally and how frequently. Some components of the interview can be missing which raises the issue of when to weight for non-response and when to impute. Alternative instruments have been developed to serve as imputation tools. Data swapping had to be done to ensure confidentiality of data from the different respondents for the same unit of analysis. Several different levels of swapping were investigated. The paper will present an overview of the many challenges that have been faced throughout the development of the survey.*

Discussant: **TBN** ???

Address: ???

Phone: ???
Fax: ???
Email: ???
List of speakers who are nonmembers: None

25.2 ssc.02

Sponsoring Society: SSC; Co-sponsored by ASA Section Stat and Environment

Session Slot: 10:30-12:20 Monday
Estimated Audience Size: ???
AudioVisual Request: ???

Session Title: Data Assimilation and Reconstruction in Canadian Environmental Research

Theme Session: No
Applied Session: Yes

Session Organizer: **Zwiers, Francis W.** Canadian Ctr for Climate Modelling & Analysis

Address: Canadian Ctr for Climate Modelling & Analysis P.O. Box 1700, Victoria, BC V8W 2Y2, CANADA
Phone: 250-363-8229
Fax: 250-363-8247
Email: fzwiers@ec.gc.ca

Session Timing: 110 minutes total (Sorry about format):

> Opening Remarks by Chair - 5 First Speaker - 30 minutes Second Speaker - 30 minutes Third Speaker - 30 minutes Discussant - 10 minutes Floor Discusion - 5 minutes

Session Chair: **Zwiers, Francis W.** Canadian Ctr for Climate Modelling & Analysis

Address: Canadian Ctr for Climate Modelling & Analysis P.O. Box 1700, Victoria, BC V8W 2Y2, CANADA
Phone: 250-363-8229
Fax: 250-363-8247
Email: fzwiers@ec.gc.ca

1. Assimilation of Data into Dynamical Models of the Coastal Ocean

Thompson, Keith R., Dalhousie University

Address: Department of Oceanography Dalhousie University Halifax, Nova Scotia Canada B3H 4J1
Phone: 902 494 3491
Fax: 902 494 2885
Email: keith@phys.ocean.dal.ca

Abstract: Driven by the need to forecast the state of the ocean and to handle the vast amount of data that streams in continuously from satellites in earth orbit, oceanographers are developing ways of assimilating data into dynamical models of the ocean. I will review some of the assimilation techniques currently in use and relate them to concepts more familiar to statisticians such as multiple regression. I will illustrate one of the techniques, the so-called adjoint method of data assimilation, by applying it to the problem of predicting tides in the Yellow Sea from coastal sea-level records. In particular I will show how this technique can be used to (i) estimate optimal open boundary conditions for a limited area model and (ii) detect errors in the water depths upon which the model is based. I will conclude by describing an extension of the readily-implemented nudging technique which, in its simplest form, involves the addition of a relaxation term to the model equations that is proportional to the difference between observed and modelled quantities.

2. Background Error Statistics Modelling in a 3D Variational Data Assimilation Scheme: Estimation and Impact On the Resulting Weather Forecasts

Gauthier, Pierre, Environment Canada

Address: Meteorological Research Branch Environment Canada 2121 Trans-Canada Highway (room 502) Dorval, P.Q. CANADA H9P 1J3

Phone: 514-421-4695

Fax: 514-421-2106

Email: pierre.gauthier@ec.gc.ca

Abstract: Since June 1997, the Canadian Meteorological Centre has replaced its previous optimal interpolation (OI) scheme by a 3D variational data assimilation (3D-var) for its global analysis. The regional analysis is also done with the 3D-var since September 1997. Although both OI and 3D-var produce a minimum variance estimate, 3D-var is a better framework to assimilate observations that are indirectly linked to the model variables such as radiance measurements. Moreover, the 3D-var implicitly uses all observations to produce the analyzed value at a grid point while OI could only used a limited number of observations. No data selection algorithm is therefore used in the variational analysis. The 3D-var also offers the possibility to use a wider variety of background error covariance models. Typically, background error correlations are considered to be homogeneous, isotropic and separable and these assumptions yield a compact representation in spectral space. Recently, experiments have been carried out to study the impact of non-separable correlations on the analysis which results in sharper wind and temperature increments. Some avenues are also explored to formulate correlation models based on empirical orthogonal functions that would relax the homogeneity and isotropy assumptions. For all correlation models, the issue of accuracy of the estimation has to be raised especially when the estimates are obtained through a time series of lagged forecasts. Perspectives are also given on how the 3D-var is a necessary step towards a 4D data assimilation scheme.

3. Optimal Estimation of Climate Parameters and Climate Data Reconstruction

Shen, Samuel S. , University of Alberta

Address: Dept. of Mathematical Sci. University of Alberta Edmonton, Alberta Canada T6G 2G1

Phone: 403-492-0216
Fax: 403-492-6826
Email: sam.shen@ualberta.ca

Abstract: *This talk discusses two optimal assessments of a climate state: optimal averaging and optimal gridding of the historical climate data.*

Spectral approach to optimal averaging: An optimal scheme is developed that minimizes the mean square error when using finitely many surface stations to measure the various orders of spherical harmonic components of a climate field. An important formula is derived to demonstrate that the sampling error is relatively insensitive to the exact shapes of empirical orthogonal functions. Two examples are given: the global average of the annual surface air temperature using 63 stations and the regional average of the monthly tropical Pacific SST.

Adaptive data gridding: The validation of climate models requires to reconstruct climate fields for the past, say 1985-1930, from the scarce observation data. A field can be reconstructed on a one-by-one degree regular grid. A systematic theory for the interpolation is described that uses the recent and more accurate observational data.

Discussant: **Switzer, Paul** Stanford University

Address: Dept. of Statistics Stanford University Stanford, CA 94305-4065 U.S.A.
Phone: 415-723-2879
Fax: 415-725-8977
Email: ps@playfair.stanford.edu

List of speakers who are nonmembers: Thompson, Shen, Gauthier

Chapter 26

ASA Journals (6)

26.1 asa.journal.jasa.app

ASA Journal Session: JASA Applications & Case Studies

Session Slot: 8:30-10:20 Tuesday
Estimated Audience Size: 100
AudioVisual Request: none

Session Title: JASA Applications & Case Studies Invited Paper

Theme Session: Yes
Applied Session: No

Session Organizer: **Lambert, Diane** Bell Labs, Lucent Technologies

Address: Bell Labs, Lucent Technologies Room 2C-256, 600 Mountain Ave. Murray Hill, NJ 07974-2008
Phone: 908 582 6509
Fax: 908 582 3340
Email: dl@bell-labs.com

Session Timing: 110 minutes total (Sorry about format):

> 110 minutes total....however you want Opening Remarks by Chair - 2 minutes First Speaker - 45 minutes Discussant - 15 minutes Discussant - 15 minutes Discussant - 15 minutes Rejoinder - 10 minutes Floor Discusion - 8 minutes)

Session Chair: **Lambert, Diane** Bell Labs, Lucent Technologies

Address: Bell Labs, Lucent Technologies Room 2C-256, 600 Mountain Ave. Murray Hill, NJ 07974-2008
Phone: 908 582 6509
Fax: 908 582 3340
Email: dl@bell-labs.com

1. Not Asked or Not Answered: Multiple Imputation for Multiple Surveys

Gelman, Andrew, Department of Statistics, Columbia University

Address: Department of Statistics, Columbia University, New York NY, 10027
Phone:
Fax:
Email: gelman@stat.columbia.edu
King, Gary, Department of Government, Harvard University
Liu, Chuanhai, Bell Labs, Lucent Technologies

Abstract: We present a way to analyze responses from a series of independent cross-sectional surveys when some questions are not asked in some surveys and some respondents do not answer all questions posed. Our method also applies to a single survey in which different respondents are asked different questions or different sampling methods are used in different strata or clusters. In all these cases, it is impossible to impute answers for each survey or respondent separately. Therefore, we combine information across surveys and respondents using a hierarchical regression model that has covariates at the levels of respondent and survey. This model allows us to impute the missing responses for items not asked or not answered. Information from survey weights is exploited in the hierarchical model by including the variables on which the survey weights were based. Population parameters of interest are estimated by reweighting responses – both observed and imputed. Comparing imputed to non-imputed data leads naturally to diagnostics for checking the fit of the imputation model. An application to pre-election public opinion polls, in which not all the

questions of interest are asked in all the surveys, that motivated the development of our model will be considered in depth.

Discussant: **Binder, David** Statistics Canada

Address: Business Survey Methods Division, 11-A R.H. Coats Building, Statistics Canada, Ottawa, Ontario, CANADA K1A 0T6
Phone: 613 951 0980
Fax: 613 951 1462
Email: binddav@statcan.ca

Discussant: **Judkins, David R.** Westat, Inc.

Address: Westat, Inc., 1650 Research Blvd. Rockville, MD 20850
Phone: 301 315 5970
Fax: 301 294 2034
Email: judkind1@westat.com

Discussant: **Santos, Robert L.** NORC

Address: 1155 East 60th St., Chicago, IL 60637
Phone: 773 256 6081
Fax: 312 753 7886
Email: judkind1@westat.com
List of speakers who are nonmembers: None

26.2 asa.journal.jasa.tm

ASA Journal Session: JASA Theory & Methods

Session Slot: 10:30-12:20 Monday
Estimated Audience Size: xx-xxx
AudioVisual Request: xxx

Session Title: JASA Theory and Methods: Invited Papers

Theme Session: No
Applied Session: Yes/No

Session Organizer: **Casella, George** Cornell University

Address: Cornell University-Ithaca 434 Warren Hall Ithaca, NY 14853-7801
Phone: 607-255-5488
Fax: 607-255-4698
Email: gc15@cornell.edu

Session Timing: 110 minutes total (Sorry about format):

> 110 minutes total....however you want Opening Remarks by Chair - 5 minutes First Speaker - 45 minutes Discussant - 15 minutes Discussant - 15 minutes Discussant - 15 minutes Floor Discusion - 15 minutes

Session Chair: **Casella, George** Cornell University

Address: Cornell University-Ithaca 434 Warren Hall Ithaca, NY 14853-7801
Phone: 607-255-5488
Fax: 607-255-4698
Email: gc15@cornell.edu

1. Approximately Exact Inference for the Common Odds Ratio in Several 2x2 Tables

Strawderman, Robert L., University of Michigan

Address: Department of Biostatistics University of Michigan 1420 Washington Heights Ann Arbor, MI 48109-2029
Phone: (734) 936-1002
Fax: (734) 763-2215
Email: strawder@umich.edu
Wells, Martin T., Cornell University

Abstract: The conditional maximum likelihood estimator of the common odds ratio in a sequence of independent 2X2 tables is known to be superior to the Mantel-Haenszel estimator in terms of asymptotic efficiency and has the further advantage that its

exact distribution is known. However, a long-standing barrier to the widespread use of this estimator has been computational intractability; in particular, the calculation of significance levels, confidence sets, and power based on the exact distribution requires fast and efficient algorithms. An important class of such algorithms form the basis of ¡italic¿StatXact¡/italic¿ (Cytel, 1992), a software package able to solve various aspects of the exact inference problem for a sequence of several 2X2 tables in real time. In this paper, we provide an alternative methodology by developing several useful saddlepoint approxmiations to the exact distribution of the conditional maximum likelihood estimator. The approximations are derived from an interesting representation for hypergeometric random variables recently developed in Kou and Ying (1996a, b), and provide fast, accurate calculations of power functions, p-values, and confidence sets. The primary computational burden is in determining the roots of a certain polynomial, which need be done numerically but only once for each table. Consequently, the required computational effort is typically minimal; for example, all of the examples in this paper were done using code written by the authors entirely in S-plus.

Discussant: **Mehta, Cyrus R.** Cytel Software Corporation

Address: Cytel Software Corporation 675 Massachusetts Ave. Cambridge, MA 02139-3309
Phone: (617) 661-4405
Fax: (617) 661-2011
Email: mehta@jimmy.harvard.edu

Discussant: **Booth, James** University of Florida

Address: Department of Statistics University of Florida Gainesville, FL 32611
Phone: (904) 392-1941
Fax: (904) 392-5175
Email: jbooth@stat.ufl.edu
Butler, Ron, Colorado State University

Address: Department of Statistics Colorado State University Fort Collins, CO 80523

Phone: (303) 491-5269
Fax: (303) 491-7895
Email: walrus@stat.colostate.edu
List of speakers who are nonmembers: None

26.3 asa.journal.tech

ASA Journal Session: Technometrics

Session Slot: 2:00- 3:50 Tuesday
Estimated Audience Size: 100-150
AudioVisual Request: xxx

Session Title: Analysis of Events Defined by Duration and Severity

Theme Session: No
Applied Session: No

Session Organizer: **Morris, Max D.** Oak Ridge National Laboratory

Address: Bldg 6012, M.S. 6367 PO Box 2008 Oak Ridge, TN 37831-6001
Phone: 615-574-3140
Fax:
Email: morris@msr.epm.ornl.gov

Session Timing: 110 minutes total (Sorry about format):

110 minutes total....however you want Opening Remarks by Chair - 5 or 0 minutes First Speaker - 35 minutes or more Discussant - 15 minutes Discussant - 15 minutes Discussant - 15 minutes Floor Discussion - 10 minutes (or 5 or 15)

Session Chair: **Morris, Max D.** Oak Ridge National Laboratory

Address: Bldg 6012, M.S. 6367 PO Box 2008 Oak Ridge, TN 37831-6001
Phone: 615-574-3140
Fax:

Email: morris@msr.epm.ornl.gov

1. Events Defined by Duration and Severity, With an Application to Network Reliability

Lambert, Diane, Bell Labs, Lucent Technologies

Address: Rm. 2C-256, 600 Mountain Ave, Murray Hill NJ 07974
Phone: 908 582 6509
Fax: 908 582 3340
Email: dl@research.bell-labs.com
Becker, Richard, A., AT&T Laboratories
Clark, Linda A., Bell Labs, Lucent Technologies

Abstract: Communications networks are highly reliable and almost never experience widespread failures. But, from time-to-time performance degrades and the probability that a call from one city to another city (say) is blocked or fails to reach its destination jumps from nearly zero to an unacceptable level. High, but variable, blocking between the two cities may then persist for a noticeable period of time. Extended periods of high blocking or events can be caused by congestion in response to natural disasters, fiber cuts, equipment failures, and software errors, for example. Because the consequences of an event depend on the level of blocking and its persistence, lists of events at specified blocking and duration thresholds, such as 50% for 30 minutes or 90% for 15 minutes, are often maintained. Reliability parameters at specified blocking and duration thresholds, such as the mean number of events per year and mean time spent in events, are estimated from the lists of reported events and used to compare network service providers, transmission facilities, or brands of equipment, for example.

This paper shows how data obtained with two-stage sampling can be used to estimate blocking probabilities as a function of time. The estimated blocking probabilities are then used to detect events and to estimate reliability parameters at specified blocking and duration thresholds. Our estimators are model-free, except for one step in a sampling bias correction, and practical even if there are

hundreds of millions of observations. Pointwise confidence intervals for reliability parameters as a function of blocking and duration thresholds are built using a kind of "partial bootstrapping" that is suitable for very large sets of data. The performance of the algorithm for event detection and the estimators of reliability parameters are explored with simulated data. An application to comparison of two network service providers is given in this paper, and possible adaptations for other monitoring problems are sketched.

Discussant: **Robinson, Jeffrey A.** General Motors R&D Center

Address: 30500 Mound Rd 1-6, Box 9055, Warren, MI 48090-9055
Phone: 810 986 1101
Fax: 810 986 0124
Email: jrobinso@rmastat.ma.gmr.com

Discussant: **Schmee, Josef** Union College

Address: 1156 Rugby Rd., Schenectady, NY 12308-2423
Phone: 518 370 6248
Fax:
Email: Jschmee@worldnet.att.net

Discussant: **TBN**

Address:
Phone:
Fax:
Email:
List of speakers who are nonmembers: None

26.4 asa.journal.jbes

ASA Journal Session: Journal of Business & Econoimc Statistics

Session Slot: 8:30-10:20 Monday
Estimated Audience Size: 80-100

AudioVisual Request: Overhead projector

Panel Session Title: Analysis of High Frequency Data: Problems and Future Research

High frequency data analysis is concerned with observations taken at small time intervals. Examples include tick-by-tick transaction data in the financial markets and the continuous ozone measurements taken by a Satellite. These data provide information on micro-structure of financial markets and ozone dynamics. They also introduce new challenges and exciting research opportunities for scientists in general and for statisticians in particular. Consider the tick-by-tick transaction data in financial markets. The sample size is huge, the transactions take place at irregular time intervals, the price may change in multiples of one-sixteenth instead of an arbitrary real number, and in many situations only the bid and ask prices are available. Traditional time series techniques are not equipped to handle such data.

The goal of this session is to ask leading researchers in four categories to provide their views and experience in analyzing high frequency data, focusing on characteristics of the data, the state of current research, and future research directions in the area. The four categories are (1) Econometrics (2) Empirical Applications in financial market micro-structure (3) Finance (4) Statistics.

Theme Session: Yes
Applied Session: Yes

Session Organizer: **Tsay, Ruey S.** University of Chicago

Address: Graduate School of Business Univ of Chicago 1101 E 58th St Chicago, IL 60637-1511
Phone: 773-702-6750
Fax: 773-702-0458
Email: rst@gsbrst.uchicago.edu

Session Organizer: **Watson, Mark** Princeton University

Address: Woodrow Wilson School Princeton University Princeton, NJ 08544

Phone: 609-258-4811
Fax: 609-258-2809
Email: mwatson@princeton.edu

Session Timing: 110 minutes total (Sorry about format):

110 minutes total Opening Remarks by Chair - 5 minutes Panelist – 20 minutes each Floor Discusion - 25 minutes

Session Chair: **Tsay, Ruey S.** University of Chicago

Address: Graduate School of Business Univ of Chicago 1101 E 58th St Chicago, IL 60637-1511
Phone: 773-702-6750
Fax: 773-702-0458
Email: rst@gsbrst.uchicago.edu
Andersen, Torben G. Northwestern University

Address: Department of Finance J.L. Kellogg Gradudate School of Mgmt Northwestern University Evanston, IL 60208
Phone: 847-467-1285
Fax: 847-491-5719
Email: t-andersen@nwu.edu
Ghysels, Eric Penn State Univ

Address: Department of Economics Penn State University 523 Kern Graduate Building University Park, PA 16802-3305
Phone: 814-865-7924
Fax: 814-863-4775
Email: eghysels@psu.edu
Polson, Nicholas Univ. of Chicago

Address: Graduate School of Business Univ of Chicago 1101 E 58th St Chicago, IL 60637-1511
Phone: 773-702-6750
Fax: 773-702-0458
Email: nick.polson@gsb.uchicago.edu
Wood, Robert Memphis State Univ

Address: Institute for the Study of Security Markets Memphis State University 32 Redthorn Cove Cordova, TN 38018

Phone: 901-755-8322

Fax: 901-753-8123

Email: raw@issm1.memphis.edu

List of speakers who are nonmembers: None

26.5 asa.journal.05

ASA Journal Session: The American Statistician

Session Slot: 10:30-12:20 Wednesday

Estimated Audience Size: 250

AudioVisual Request: only standard projector and screen

Session Title: Undergraduate Statistics Education: What Should Change?

The focus of this session will be on activities within statistics departments at research universities that are focused on "improving the pipeline" into our profession. This is in contrast to the activities involving graduate training in statistics and statistical research, which are the focus of the vast majority of sessions at the JSM (and rightfully so). The Higgins paper makes the case that the undergraduate statistics major needs to take a big step away from mathematics and towards computer science in order to be a viable option in the coming years. The Hogg paper is concerned more with the process of the non-faculty-research activities that statistics departments undertake, and how ideas from CQI (Continuous Quality Improvement) can be adapted to statistics. Thus, the two papers complement each other nicely in that the Higgins paper is concerned with the CONTENT of what departments teach for the undergraduate major, while the Hogg paper is concerned with HOW the department's work is delivered, evaluated, and improved.

Theme Session: No

Applied Session: No

Session Organizer: **Landwehr, James M.** Bell Labs, Lucent Technologies

Address: Bell Labs, Lucent Technologies Room 2C-257 700 Mountain Ave. Murray Hill, NJ 07974-0636
Phone: 908-582-7405
Fax: 908-582-3340
Email: jml@bell-labs.com

Session Timing: 110 minutes total (Sorry about format):

> 110 minutes total Opening Remarks by Chair - 5 minutes First Speaker - 25 minutes Second Speaker - 25 minutes Discussant - 10 minutes Discussant - 10 minutes Discussant - 10 minutes Floor Discusion - 25 minutes

Session Chair: **Landwehr, James M.** Bell Labs, Lucent Technologies

Address: Bell Labs, Lucent Technologies Room 2C-257 700 Mountain Ave. Murray Hill, NJ 07974-0636
Phone: 908-582-7405
Fax: 908-582-3340
Email: jml@bell-labs.com

1. Nonmathematical Statistics: A New Direction for the Undergraduate Discipline

Higgins, James J., Kansas State University

Address: Department of Statistics, Dickens Hall Kansas State University Manhattan, KS 66506-0802
Phone: 785-532-0509
Fax: 785-532-7736
Email: jhiggins@ksu.edu

Abstract: The explosion in the amount of data available to society today has not led to a corresponding growth in undergraduate statistics programs and departments to meet the demand for statisticians. Instead, the profession is faced with the specter of "statistics departments under siege." For statistics to be recognized as a fundamental undergraduate discipline worthy of departmental status and continued support, it must go much further *than it has to establish its own identity. This can only be done if ties to mathematics and mathematical sciences departments are finally broken. To do this, statisticians must develop the non-mathematical side of the discipline. This is important not only for the identity of statistics but for its future as a viable undergraduate discipline. A view of what a restructured undergraduate statistics curriculum might look like is put forth. Courses are suggested that would meet important needs of the undergraduate statistics major and set the discipline of statistics apart from mathematics.*

2. Let's Use CQI In Our Statistics Programs

Hogg, Robert V., University of Iowa

Address: Department Of Statistics & Actuarial Science 214 Schaeffer Hall University of Iowa Iowa City, IA 52242
Phone: 319-335-0824
Fax: 319-335-3017
Email: bhogg@stat.uiowa.edu

Abstract: Continuous Quality Improvement in higher education is a process in which an attempt is made to optimize the quality output of an individual, a department, or a college. The optimization of any system of interdependent parts can not be achieved by optimizing the parts separately. In particular, we must consider how best to use faculty members, who have different talents, to fulfill the three parts of the mission: research/discovery, teaching/learning, and professional practice/service, eliminating waste and non-value added activities as much as possible. The remarks of this discussion are directed primarily at improving the teaching/learning part of the mission. These involve, among other items: better communications, (this means listening also, Hogg), including appropriate feedback from students and faculty, possibly having retreats; elimination of certain wasteful activities, more generally personally quality improvement of faculty and students; discussion of course experiences, in particularly, some mentoring between senior and junior faculty; serious considerations how to improve service courses, most of which are given

through large lectures; recruitment of students to undergraduate programs, which need changes to improve quality; improved relationships among the stakeholders of statistical education; involvement of students, who have a great deal of energy, in projects on campus and with partners in industry.

Discussant: **Cobb, George W.** Mt. Holyoke College

Address: Dept. of Mathematics, Statistics and Computer Science Mt. Holyoke College South Hadley, MA 01075-1411
Phone: 413-538-2401
Fax: 413-538-2239
Email: gcobb@mtholyoke.edu

Discussant: **Newton, H. Joseph** Texas A&M University

Address: Department of Statistics 447 Blocker Building Texas A&M University College Station, TX 77843-3143
Phone: 409-845-3141
Fax: 409-845-3144
Email: jnewton@stat.tamu.edu

Discussant: **Wasik, John L.** North Carolina State University

Address: Dept. of Statistics Box 8203 North Carolina State Univsity Raleigh, North Carolina 27695-8203
Phone: 919-515-1956
Fax: 919-515-7591
Email: wasik@stat.ncsu.edu
List of speakers who are nonmembers: None

Chapter 27

ASA Committees (3 or 4)

27.1 asa.committees.01

ASA COMMITTEE ON PRIVACY AND CONFIDENTIALITY

Session Slot: 10:30-12:20 Tuesday
Estimated Audience Size: 100
AudioVisual Request: regular – overhead projector

Panel Session Title: Should We Continue To Release Public-Use Microdata Files?

Statistical data from censuses and surveys are disseminated in text, tabular, graphic, and microdata formats. The last of these, which has substantially enhanced the utility of census and survey data, is a product of the computer age. Data users receive files of individual records, giving them the ability to interact with the data set and perform the tabulations and statistical analyses that are best suited to their own purposes. Two principles have been applied to the release of microdata files - restricted data and restricted access. For virtually all releases of microdata files, data are restricted by removing explicit identifiers and taking other steps to limit the possibility that individuals can be identified. In some instances, access is restricted to certain categories of users, such as those who have research grants with a federal agency or those who are willing to work with the files at a federal site. Users in these categories are generally subject to penalties for violating the terms of their access to the files. A public-use microdata file is one for which there are no restrictions on access, other than payment of the cost of a tape or CD- ROM. Some public-use microdata files can be downloaded via the Internet. In recent years, Rubin, Fienberg, and others have proposed to eliminate all risk that individuals can be identified by releasing microdata files containing only synthetic data - records that do not contain information for specific individuals but, in the aggregate, retain the statistical properties of the real records.

Several federal statistical agencies are currently releasing public-use microdata sets and plan to do so in the future, e.g., from the 2000 Census of Population. However, some respected statisticians argue that this practice cannot continue indefinitely - that with the rapid development of more efficient computer matching programs and the proliferation of computerized information about individuals available from both the public and private sectors, the risks of disclosure of the identities of persons whose records are included in public-use data sets will become unacceptably high. The purpose of this panel discussion is to discuss the risks and benefits associated with continued release of public-use microdata files and to discuss the pros and cons of alternative approaches - restricted access and release of synthetic data files.

Theme Session: Yes
Applied Session: Yes

Panel Organizer: **Jabine, Thomas B.** Independent Consultant

Address: 3231 Worthington St., NW, Washington, DC 20015-2362
Phone: 202-244-4179
Fax:
Email: tjabine@nas.edu

Panel Organizer: **de Wolf, Virginia A.** US Bureau of Labor Statistics

Address: 1719 Luzerne Avenue, Silver Spring, MD 20910-1510
Phone: 202-606-7374
Fax: 202-606-7426
Email: dewolf_v@bls.gov

Session Timing: 110 minutes total (Sorry about format):

110 minutes total...please allocate Opening Remarks by Chair - 5 minutes First Panelist - 15 minutes Second Panelist - 15 minutes Third Panelist - 15 minutes Fourth Panelist - 15 minutes Fifth Panelist - 15 minutes Rebuttals/Discussion Among Panel Members - 15 minutes Floor Discussion - 15 minutes

Panel Chair: **Kirkendall, Nancy J.** US Office of Management and Budget

Address: OMB, OIRA, Statistical Policy Branch, NEOB, Room 10201, 725-17th St., NW, Washington, DC 20503
Phone: 202-395-7315
Fax: 202-395-7245
Email: nancy_kirkendall@oa.eop.gov
David, Martin University of Wisconsin

Address: Univ. of Wisc., Dept. of Economics, 6440 Soc. Sci., Bldg., 1180 Observatory Dr., Madison, WI 53706
Phone: 608-262-3281
Fax: 608-263-3876
Email: david@ssc.wisc.edu
Fienberg, Stephen E. Carnegie Mellon University

Address: Carnegie Mellon University, Dept. of Statistics, Pittsburgh, PA 15213
Phone: 412-268-2723
Fax: 412-268-7828
Email: fienberg@stat.cmu.edu
Gordon, Nancy M. Census Bureau

Address: Census Bureau, Room 2061-3, Washington, DC 20233
Phone: 301-457-2126
Fax: 301-457-3761
Email: nancy.m.gordon@ccmail.census.gov
Juster, F. Thomas University of Michigan

Address: Univ. of Michigan, Inst. for Social Research, PO Box 1248, Room 3240, Ann Arbor, MI 48106-1248
Phone: 313-764-4207

Fax: 313-647-1186
Email: ftjuster@umich.edu
Scheuren, Frederick J. Ernst & Young

Address: 1402 Ruffner Road, Alexandria, VA 22302-4217
Phone: 202-327-6000
Fax:
Email: fritz.scheuren@ey.com
List of speakers who are nonmembers: None

27.2 asa.committees.02

MINORITIES IN STATISTICS COMMITTEE. Co-sponsor: Statistics in Epidemiology Section

Session Slot: 2:00- 3:50 Wednesday
Estimated Audience Size: 65
AudioVisual Request: Slide/Overhead Projector

Panel Session Title: Are We Getting the Right Answers; Are We Asking the Right Questions?

There are huge disparities in health in between minorities and nonminorities. We need to ask why these differences exist. We need to develop prevention and intervention strategies appropriate for the different cultures. If we are to have solutions that are realistic and effective in different cultures, then we will need the unique insight and experience of minority researchers and community workers to ask the right questions, to ask the questions in the right way, to observe and interpret and to draw conclusions about and make recommendations for people and cultures with completely different life experiences and life circumstances.

This session will bring persons from different racial and ethnic groups to discuss our paradigms and how they affect our ability to ask the right questions and get more realistic and effective answers/solutions in diverse cultures.

Theme Session: Yes

Applied Session: Yes
Panel Organizer: **Reynolds, Gladys H.** CDC

Address: Office of the Director, MS D39 Centers for Disease Control and Prevention 1600 Clifton Rd., NE Atlanta, GA 30333
Phone: (404) 639-7224
Fax: (404) 639-7039
Email: GHR1@CDC.GOV

Session Timing: 110 minutes total (Sorry about format):

NA

Panel Chair: **Aliaga, Martha** University of Michigan

Address: Statistics Department University of Michigan 1450-A Mason Hall Ann Arbor, MI. 48109
Phone: (313) 764-5493
Fax: (313) 763-4676
Email: aliaga@umich.edu
Reynolds, Gladys H. Centers for Disease Control and Prevention

Address: Office of the Director, MS D39 Centers for Disease Control and Prevention 1600 Clifton Rd., NE Atlanta, GA 30333
Phone: (404) 639-7224
Fax: (404) 639-7039
Email: GHR1@CDC.GOV
McPheron-Alex, Theda West Virginia University

Address: West Virginia University 126 Prairie Avenue Morgantown, WV 26505
Phone: (304) 291-5126
Fax: (304) 291-5126
Email: anasazi@access.mountain.net
Jenkins, Bill Centers for Disease Control and Prevention

Address: National Center for HIV, STD and TB Prevention (MS-E02) Centers for Disease Control and Prevention 1600 Clifton Road Atlanta, Ga. 30333

Phone: (404) 639-8367
Fax: 404-639-8629
Email: WCJ1@CDC.GOV
Santos, Robert L. University of Chicago

Address: University of Chicago 1155 E. 60th St. Chicago, Ill 60637
Phone: (773) 256-6081,
Fax: (773) 753-7540
Email: santosro@norcmail.uchicago.edu
Liao, Youlian Loyola U Medical Center

Address: Biostatistics Section Dept. of Preventive Medicine Loyola U Medical Center 2160 South First Avenue Maywood, Ill. 60153
Phone: 708 327-9015
Fax: 708 327 -9009
Email: yliao@maya.medctr.luc.edu
List of speakers who are nonmembers: McPheron,Liao

27.3 asa.committees.03

Committee for Applied Statisticians

Session Slot: 2:00- 3:50 Tuesday
Estimated Audience Size: 200
AudioVisual Request: xxx

Panel Session Title: Doing Statistics in a Non-Statistical World: Common Challenges Facing Applied Statisticians

Theme Session: Yes
Applied Session: Yes
Panel Organizer: **Warshaw, Meredith** Brown University

Address:
Phone:
Fax:
Email:
Panel Organizer: **Cole, Jonathan** New York State Department of Labor

Address: 4 Martin Lane, Troy, NY 12182
Phone: 518-457-1354
Fax: 518-485-7810
Email: joncole@global2000.net

Session Timing: 110 minutes total (Sorry about format):

Panel

Panel Chair: **Cole, Jonathan** New York State Department of Labor

Address: 4 Martin Lane, Troy, NY 12182
Phone: 518-457-1354
Fax: 518-485-7810
Email: joncole@global2000.net
Klemm, Rebecca Klemm Analysis Group Inc

Address: 1785 Massachusetts Avenue, Washington, DC 20036
Phone: 202-667-5244
Fax: 202-667-5793
Email: rklemm@klemmanalysis.com
Iman, Ron Southwest Technology Consultants

Address: 1065 Trammway Lane NE, Albuquerque, NM 87122
Phone: 505-856-6500
Fax: 505-856-1300
Email: ron_iman@compuserve.com
Karen Pieper Pieper Statistical Consulting

Address: 32 Ivy Trails, Weaverville, NC 28787
Phone: 704-649-0149
Fax: 704-649-0230
Email: kpieper@mindspring.com
Bergum, James Bristol-Myers Squibb

Address: 1 Squibb Drive PO Box 1, New Brusnwick, NJ 08903-0191
Phone: 732-519-3517
Fax: 732-519-2351
Email: james_s._begum@ccmail.bms.com
List of speakers who are nonmembers: None

27.4 asa.committees.04

Sponsoring Section/Society: ASA COMMITTEE ON WOMEN IN STATISTICS, CAUCUS FOR WOMEN IN STATISTICS

Session Slot: 2:00- 3:50 Monday
Estimated Audience Size: 100
AudioVisual Request: xxx

Session Title: The Role and Status of Women in Statistics Profession

In this session, opportunities and challenges for female statisticians will be discussed. The speakers will examine the issue from the aspects of academia, industry and government. Dr. Olkin will begin the presentation from the academia point of view. He will review the problems with evaluation procedures, the shortage of tenured women faculty in statistics department, and why women aren*t chosen for key jobs, etc. Ms. Chen will present the ratio of male/female statisticians and other relevant statistics in the industrial setting, followed by the discussions centering in the pharmaceutical industry, especially in the contract research organization (CRO). Specific issues for female statisticians in the environment will be discussed. Dr. Clark will show that in the U. S. Federal Government, the skills and efforts of women statisticians have been well recognized and therefore, the opportunities for women statisticians have been greatly expanded. She will investigate how women have benefited from the changes and have become leaders in the federal statistical agencies. Specific examples will be given. At the end of the session, Dr. Billard will summarize the three presentations and give an insight discussion on some specific issues.

Theme Session: Yes
Applied Session: Yes

Session Organizer: **Ju, Huey Lin** National Chengchi University/ Visiting Stanford University 1997/98

Address: 1189 Fynes Ct., San Jose, CA 95131
Phone: (408)4375772
Fax: (408)4375772

Email: hju@stat.stanford.edu

Session Timing: 110 minutes total (Sorry about format):

110 minutes total to be split up Opening Remarks by Chair - 0 First Speaker - 30 minutes Second Speaker -30 minutes Third Speaker - 30 minutes Discussant - 15 minutes (or none) Floor Discussion - 5 minutes

Session Chair: **Ju, Huey Lin** National Chengchi University/ Visiting Stanford University 1997/98

Address: 1189 Fynes Ct., San Jose, CA 95131
Phone: (408)4375772
Fax: (408)4375772
Email: hju@stat.stanford.edu

1. The Status of Women in Academic Statistics

Olkin, Ingram, Stanford University

Address: Department of Statistics Sequoia Hall Stanford University Stanford, CA 94305-4065
Phone: (650 7252245
Fax: (650)7258977
Email: iolkin@stat.stanford.edu

Abstract: We review the status of women in academia in therms of tennured faculty and in terms of involvement in societal activities, and some aspects of a mentoring program. A probabilistic analysis of the mechanism by which candidates are chosen on shortlists of candidates will also be presented.

2. Opportunities and Challenges for Women in the Pharmaceutical (CRO) Industry

Chen, Mon-Gy, Covance, Inc.

Address: 210 Carnegie Center, Princeton, NJ 08540
Phone: (609)4524564
Fax: (609)9879048

Abstract: *In the past decade, the pharmaceutical industry has undergone many changes such as reengineering due to health care reform. These changes offer Contract Research Organization (CRO) great opportunities and challenges in drug research and development. The primary responsibility of CRF is to provide pharmaceutical companies technical expertise and efficient project management to speed up the drug development and approval. In this presentation, the opportunities and challenges of female statisticians in pharmaceutical industry, especially the CRO industry, will be discussed.*

3. Career Opportunities for Statisticians in Federal Statistical Agencies

Clark, Cynthia Z.F., Bureau of the Census

Address: Methodology and Standard Bureau of the Census Washington, D.C. 20233

Phone: (301) 4572160

Fax: (301)4571902

Email: Cynthia_Z_f_clark@Census.Gov

Abstract: *Across most disciplines and job classifications, the U.S. Federal Government has played a lead role in expanding opportunities for all individuals. Managers are required to consider fairly all individuals in all work settings. Women have particularly benefitted from this workforce policy. The opportunities for women statisticians and their influence expanded as their diverse skills became recognized in the workplace. Women have risen to the occasion and have become leaders in the Federal Statistical Agencies. This new management diversity continues to enhance the career development opportunities for both women and men in government statistics. This paper will provide specific examples of these opportunities.*

Discussant: **Billard, Lynne** University of Georgia

Address: Department of Statistics University of Georgia Athens, GA 30602-1952

Phone: (706)5423281
Fax: (706)5423391
Email: Lynne@stat.uga.edu
List of speakers who are nonmembers: None

Chapter 28

ASA Chapters (2)

28.1 asa.chapters.01

Sponsoring Section/Society: ASA Council of Chapters

Session Slot: 8:30-10:20 Thursday
Estimated Audience Size: 80
AudioVisual Request: xxx

Session Title: MCMC Methods in Practice

Theme Session: Yes/No
Applied Session: Yes

Session Organizer: **Dey, Dipak K.** University of Connecticut

Address: Department of Statistics, Box U-120, Storrs, Ct 06269-3120
Phone: (860)486-4196
Fax: (860)486-4113
Email: dey@stat.uconn.edu

Session Timing: 110 minutes total (Sorry about format):

110 minutes total...please allocate Opening Remarks by Chair - 5 First Speaker - 30 minutes Second Speaker - 30 minutes Third Speaker - 30 minutes

Floor Discussion - 15 minutes

Session Chair: **Dey, Dipak K.** University of Connecticut

Address: Department of Statistics, Box U-120, Storrs, Ct 06269-3120
Phone: (860)486-4196
Fax: (860)486-4113
Email: dey@stat.uconn.edu

1. Assessing Quality of Care Following a Heart Attack Using Monte Carlo Markov Chain Methods

Normand, Sharon-Lise, Harvard School of Public Health, Boston

Address: Department of Health Care Policy
Phone: (617)432-3260
Fax: (617)432-0173
Email: sharon@figaro.med.harvard.edu
Landrum, Mary Beth, Harvard Medical School

Abstract: Medical practice guidelines are increasingly utilized to understand variations in the delivery of many medical and surgical procedures. The most common approach to developing guidelines is to elicit judgements from a multi-disciplinary panel of experts regarding treatment efficacy within distinct clinical strata. In this talk, we employ Monte Carlo Markov Chain methods to develop a standard of care for patients who have suffered a heart attack and subsequently, profile the quality of care delivered to a sample of Medicare patients who were treated in US hospitals during 1994 - 1995.

2. Analyzing Real Estate Data Problems Using the Gibbs Sampler

Ghosh, Sujit K., North Carolina State University, Raleigh

Address: Department of Statistics, Raleigh, NC 27695
Phone: (919)515-1950
Fax: (919)515-7591
Email: sghosh@stat.ncsu.edu
Gelfand, Alan E., University of Connecticut
Knight, John R., Eberhardt School of Business, University of Pacific

Sirmans, C.F., University of Connecticut

Abstract: Real estate data is often characterized by data irregularities, e.g., missing data, censoring or truncation, measurement error, etc. Practitioners often discard missing or censored data cases and ignore measurement error concerns. We argue here that we can remedy these irregularity problems through simulation based model fitting using the Gibbs sampler. We describe the Gibbs sampler in the context of these issues focusing primarily on the missing data problem. We illustrate using a sample of residential property sales from Baton Rouge, Louisiana. We show dramatic improvement in inference by retaining the partially observed data cases rather than deleting them. We also detail how the other problems can be handled using the Gibbs sampler. We conclude that, while canned *software to implement a Gibbs sampler does not exist, for the problems at hand, development is straightforward with substantial reward anticipated for the effort.*

3. Bayesian Inference for Estimating Hunting Success Rates Based on Survey Data

He, Zhuoqiong, Missouri Department of Conservation

Address: 1110 S College Avenue, Columbia, MO 65201
Phone: (314)882-9880
Fax: (314)874-8849
Email: zqhe@statg.cs.missouri.edu
Sun, Dongchu, University of Missouri, Columbia

Abstract: Although most post-season harvest surveys are conducted at state level, the effective management of wildlife population often wants to estimate hunting success rates, hunting pressure and harvest at sub-area (such as management unit, region, or county) level. The sample sizes for some sub-areas are often very small or even zero. It is well known that direct survey estimates for small sub-areas are often yield unacceptably large standard errors due to small sample sizes. In this article, a Bayesian

hierarchical linear mixed model is used to estimate hunting success rates at sub-area level for post season harvest surveys. The model includes fixed geographic variables such as forest coverage, random geographic variables such as regional differences, longitudinal variables such as time effects, and spatial variables such as the correlation between neighboring sub-areas. Markov chain Monte Carlo methods such as Gibbs sampling and adaptive rejection sampling are used to compute the Bayesian estimators of the parameters. The method is illustrated using data from Missouri Turkey Hunting Survey in the Spring of 1996. The Bayesian estimates are close to the frequency estimates for the counties with large sample sizes and quite stable for those counties with small sample sizes. Bayesian model selection is used to show that the spatial correlation between counties does exist.

List of speakers who are nonmembers: None

28.2 asa.chapters.02

Sponsoring Section/Society: North Texas Chapter of the ASA

Session Slot: 2:00- 3:50 Sunday
Estimated Audience Size: 200
AudioVisual Request: xxx

Session Title: Spatial-Temporal Modeling of Environmental Processes

Theme Session: No
Applied Session: No

Session Organizer: **Gunst, Richard F.** Southern Methodist University

Address: Department of Statistical Science Southern Methodist University Dallas, TX 75275-0332
Phone: 214-768-2466
Fax: 214-768-4035
Email: rgunst@mail.smu.edu

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 5 minutes First Speaker - 30 minutes Second Speaker - 30 minutes Third Speaker - 30 minutes Discussant - 0 minutes Floor Discusion - 15 minutes

Session Chair: **Gunst, Richard F.** Southern Methodist University

Address: Dept. of Statistical Sci. Southern Methodist University Dallas, TX 75275-0332
Phone: 214-768-2466
Fax: 214-768-4035
Email: rgunst@mail.smu.edu

1. Modeling Contamination Migration with Spatially Correlated Stochastic Differential Equations

Hartfield, Molly, Radian International

Address: Radian International P.O. Box 201088 Austin, TX 78720-1088
Phone: 512-419-6241
Fax: 512-454-8807
Email: Molly_Hartfield@radian.com

Abstract: A class of spatially correlated stochastic differential equation models are applied to the problem of modeling space-time data. In particular, the models are applied to environmental data, where interest is on monitoring contaminant levels in the soil or groundwater, characterizing the rate at which contaminant concentrations are changing in time, or tracking the migration of contaminants across a region. The class of models can be described as continuous-time versions of spatial autoregressive integrated stochastic processes. Model identification techniques borrow from spatial statistical model identification methods including sample semivariogram plots and plots based on generalized increments. Because these spatial statistical techniques were originally developed to accommodate the irregularity that often characterizes spatial data, the model identification methods for the space-time models considered here also do not require that data are collected at evenly spaced time intervals or along a regular spatial grid. An *recursive scheme for fully specifying a space-time model and estimating its parameters is applied. Restricted maximum likelihood (REML) methods are used to estimate model parameters and the Kalman filter is demonstrated to be a convenient computational tool for obtaining the REML estimates and providing model diagnostics.*

2. Criteria Air Pollutant Levels in Space and Time

Ensor, Katherine Bennett, Rice University

Address: Department of Statistics, MS 138 Rice University Houston, TX 77251-1892
Phone: 713-527-4687
Fax: 713-285-5476
Email: kathy@stat.rice.edu
Calizzi, Mary, Rice University
Baggett, Scott, Rice University

Abstract: A state-space formulation of a multivariate spatial-temporal model is developed and applied to air quality data for the Houston area. The goal is to provide an empirical method for predicting the level of ambient pollutants such as ozone, nitric oxides and small particulates for the study area. The major physical components of the system under study, e.g. the pollutant transport due to wind and ozone formulation due to favorable meteorological conditions, are included in the modeling effort.

3. Samping Errors in Some Global Climate Sampling Schemes

Li, Ta-Hsin, University of California at Santa Barbara

Address: Department of Statistics University of California Santa Barbara, CA 93106-3110
Phone: 805-893-4760
Fax: 805-893-2334
Email: thl@pstat.ucsb.edu

Abstract: Spherical harmonics are employed in many general circulation models (GCMs) as part of their numerical scheme. There arises the problem of estimating the coefficients in the spherical harmonic expansion from imperfect observing systems. For climate change assessments, it is important to understand the uncertainty in the estimation of the coefficients that represent the amplitude of spherical waves.

Sampling errors are inevitable when the spherical harmonic coefficients are estimated from observations made by a discrete set of monitoring stations or by asynchronous satellites. The sampling errors not only result from a lack of sufficient density in the monitoring stations and the satellite orbits, but also result, for satellite data, from the time span required for an asynchronous satellite to cover the globe. The sampling errors are menifested as aliasing—the tendency for variations in unresolved scales to contaminate the estimates in resolved scales.

A methodology for systematic analysis of sampling errors is provided. Based on a simple noise-forced energy balance model (EBM), the sampling errors are investigated both analytically and numerically for some sampling schemes. Special attention is paid to the Gauss-Legendre design in which the sampling points are equally spaced along latitude circles but located at the zeros of Legendre polynomial. For the satellte sampling, polar-orbiting satellites are of special interest because of their widespread use in operational monitoring of the climate system. In this study, the Gauss-Legendre design is found to be superior to the latitude-longitude uniform design because it produces smaller errors given the same number of sampling points. For polar-orbiting satellites, the samping errors are found to depend crucially on the time scale of the climate field being sampled—the faster the field changes with time the larger the sampling errors.

List of speakers who are nonmembers: None

Chapter 29

Other Associations (6 — note number 3 moved)

29.1 asa.other.ass.01

International Indian Statistical Association

Session Slot: 8:30-10:20 Thursday
Estimated Audience Size: 40-60
AudioVisual Request: xxx

Session Title: Inference For Stochastic Processes

This topic provides an interface between stochastic processes and statistical inference. Most stochastic models depend on unknown parameters or other nonparametric quantities such as unknown densities etc. Statistical inference from dependent data is an important and a rapidly growing area of research. This session contains three talks on inference for nonlinear spatial data (S. Lele), minimum Hellinger distance estimation for branching processes (T.N. Sriram) and filtering via estimating functions (A. Thavaneswaran and M.E. Thompson).

Theme Session: No
Applied Session: Yes

Session Organizer: **Basawa, Ishwar** University of Georgia

Address: Department of Statistics University of Georgia Athens,GA 30602

Phone: 706-542-3309
Fax: 706-542-7069
Email: ishwar@stat.uga.edu

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 5 or 0 minutes First Speaker - 30 minutes (or 25) Second Speaker - 30 minutes Third Speaker - 30 minutes Discussant - 10 minutes (or none) Floor Discusion - 10 minutes (or 5 or 15)

Session Chair: **Thompson, Mary E.** University of Waterloo

Address: Statistics Dept. University of Waterloo,Waterloo ON Canada N2L 3G1
Phone: 519-885-1211 Ext. 5543
Fax:
Email: methomps@setosa.uwaterloo.ca

1. Modeling and Inference for Nonstationary Spatial Data

Lele, Subhash, Johns Hopkins University

Address: Department of Biostatistics Johns Hopkins University Baltimore, MD 21205
Phone:
Fax:
Email: slele@welchlink.welch.jhu.edu

Abstract: In modeling and inference for spatial data, it is commonly assumed that the covariance structure is stationary and isotropic, meaning that the covariance between two locations depends only on the distance between them. The non-stationarity in the process is usually modelled through trend or dependence on the spatial covariates. However, as in regression, we seldom know or are aware of all the covariates that should be put into *the model. The missing covariates in the trend introduce nonstationarity in the covariance parameter. In spatial statistics, it is commonly assumed that the effect of the missing covariates can be accounted for by the use of the dependence or the neighboring values. Although this is true for the point prediction, it is not true when it comes to coverage properties of the prediction intervals. We introduce a hierarchical model for modelling nonstationarity in the covariances. We use the Monte-Carlo Newton Raphson algorithm for estimation of the parameters. We show how prediction intervals for an observation at a new location can be obtained under this model. The prediction intervals obtained under the hierarchical model are shown to have better coverage properties than the prediction intervals obtained under stationary, isotropic model fitted with the same misspecified trend as the hierarchical model.*

2. Minimum Hellinger Distance Eestimation for Supercritical Branching Orocesses

Sriram, T.N., University of Georgia

Address: Department of Statistics, University of Georgia Athens, GA 30602

Phone: 706-542-3305

Fax: 706-542-3391

Email: tn@stat.uga.edu

Vidyasankar, A., University of Georgia

Abstract: *For supercritical branching processes, suppose we postulate that the probability mass function of the offspring variable belongs to a specified parametric family* $\{f_\theta; \theta \in \Theta\}$*. If the postulated model is correct, then one can use the MLE* $\hat{\theta}_{MLE}$ *to estimate* θ *and it is known that* $\hat{\theta}_{MLE}$ *is asymptotically efficient. However, it is also well known that MLEs do not, in general, possess the property of stability under small perturbations in the underlying model. To overcome this deficiency, we propose a minimum Hellinger distance estimator* $\hat{\theta}_{MHD}$*. We establish its existence, uniqueness, consistency and derive the limit distribution. It is shown that (a)* $\hat{\theta}_{MHD}$ *is asymptotically efficient (just as the*

MLE is) if the postulated model is in fact true and (b) the limit distribution of $\hat{\theta}_{MHD}$ is not greatly perturbed if the assumed model is only approximately true. Robustness properties pf $\hat{\theta}_{MHD}$ are also studied.

3. Recent Developments in Filtering Via Estimating Functions

Thavaneswaran, A., Department of Statistics University of Manitoba

Address: University of Manitoba Winnipeg Manitoba R3T 2N2, Canada
Phone: 204-474-8984
Fax: 204-474-7621
Email: thavane@ccm.umanitoba.ca
Thompson, Mary E., University of Waterloo

Abstract: Godambe's (1995) theorem is applied to continuous time stochastic processes. Filtering formulas are given for short memory semimartingale models as well as for long memory self similar processes. Superiority of the approach over Kalman filtering with non-Gaussian processes is also discussed in some detail.

List of speakers who are nonmembers: None

29.2 asa.other.ass.02

National Institute of Statistical Sciences

Session Slot: 10:30-12:20 Wednesday
Estimated Audience Size: xx-xxx
AudioVisual Request: xxx

Session Title: Theoretical and Methodological Issues Arising from NISS Research

.

Theme Session: Yes/No
Applied Session: Yes/No

Session Organizer: **Karr, Alan F.** National Institute of Statistical Sciences

Address:
Phone:
Fax:
Email:

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 5 or 0 minutes First Speaker - 30 minutes (or 25) Second Speaker - 30 minutes Third Speaker - 30 minutes Discussant - 10 minutes (or none) Floor Discusion - 10 minutes (or 5 or 15)

Session Chair: **Karr, Alan F.** National Institute of Statistical Sciences

Address:
Phone:
Fax:
Email:

1. Statiatical Questions for Freeway Traffic: Travel Time Estimation and Prediction, Incident Detection and Origin/Destination Tables

Bickel, Peter J., University of California Berkeley

Address:
Phone:
Fax:
Email:
Rice, John, University of California Berkeley
Ritov, Yaacov, Hebrew University of Jerusalem
Petty, Karl, University of California Berkeley

Abstract:

2. ***Combining Toxicological Information***
Simpson, Douglas, University of Illinois at Urbana/Champaign

Address:
Phone:
Fax:
Email:
Carroll, Raymond J., Texas A&M University
Xie, Minge, Rutgers University

Abstract:

3. ***Reconstruction of Contingency Tables***
West, Mike, Duke University

Address:
Phone:
Fax:
Email:
Tebadi, Claudia, NCAR

Abstract:

List of speakers who are nonmembers: None

29.3 asa.other.ass.04

Sponsoring Section/Society: Caucus for Women in Statistics

Session Slot: 2:00- 3:50 Tuesday
Estimated Audience Size: 150
AudioVisual Request: Overhead Projector, Slide Projector

Session Title: Statistical Models for Individualized Counseling about Breast Cancer

Theme Session: Yes
Applied Session: No

Session Organizer: **Parmigiani, Giovanni** Duke University

Address: Giovanni Parmigiani Box 90251, ISDS Duke University Durham, NC 27708-0251
Phone: 919-684-8064
Fax: 919-684-8594
Email: gp@stat.duke.edu

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 5 minutes First Speaker - 30 minutes Second Speaker - 30 minutes Third Speaker - 30 minutes Discussant - 0 minutes Floor Discusion - 15 minutes

Session Chair: **Stangl, Dalene** Duke University

Address: Dalene Stangl Box 90251, ISDS Duke University Durham, NC 27708-0251
Phone: 919-684-4263
Fax: 919-684-8594
Email: dalene@stat.duke.edu

1. Predicting Risk of Breast Cancer

Benichou, Jacques, Centre Hospitalier Universitaire de Rouen, France

Address: Jacques Benichou, M.D., Ph.D. Division of Cancer Epidemiology and Genetics National Cancer Institute Executive Plaza North, Room 403 6130 Executive Blvd. MSC 7368 Bethesda, MD 20892-7368
Phone: 301-496-3356
Fax: 301-402-0081
Email: Jacques_Benichou@nih.gov

Abstract: *When counseling women about their breast cancer risk, it is useful to provide a quantitative assessment of their individualized risk of developing breast cancer, or absolute risk. Absolute risk is defined as the probability that a woman with a given risk profile and given age will develop breast cancer over a defined time period.*

There are three general approaches to estimate absolute risk of breast cancer depending on how familial and genetic risk factors are taken into account. In the empirical approach, observable family history as well as other non familial risk factors are modeled and no genetic model is assumed nor any genotype observed. An intermediate or genetic-modeling approach consists in analyzing family history data relying on a genetic model, but genotype is again not observed. Finally, in the direct approach, genotype is directly measured and modeled as a risk factor.

In this paper, the principles, results, limitations and possible extensions of these three approaches are reviewed.

The empirical approach has been implemented in the Gail et al. model. Based on the data from the Breast Cancer Demonstration and Detection Project (BCDDP), absolute risk estimates of breast cancer were derived by combining data on the five-year follow-up of 243,221 white women and data from a nested case-control sample of 2,852 cases and 3,146 controls. Absolute risk estimates depend on current age as well as four risk factors, namely family history of breast cancer in first-degree relatives, age at menarche, age at first live birth and number of previous breast biopsies. A computer program and graphs allow counselors to easily implement this approach and provide point estimates as well as confidence intervals for the absolute risk of breast cancer. Besides individual counseling, these absolute risk predictions have been used to define inclusion criteria and calculate sample size in a trial to determine whether Tamoxifen can reduce the risk of developing breast cancer. Validation studies showed that this model provides accurate predictions for women in regular screening but tends to exaggerate the risk for unscreened or sporadically screened women.

2. Family History of Breast Cancer: Controlling for the Effect

of BRCA1/BRCA2

Claus, Elizabeth B., Yale University

Address: Elizabeth Claus Department of Epidemiology & Public Health Yale University School of Medicine 60 College Street P.O. Box 20-8034 New Haven, Connecticut 06520-8034

Phone: 203-785-6050

Fax: 203-785-6912

Email: elizabeth.claus@yale.edu

Abstract: Approximately ten percent of women in the Unites States have a family history of breast cancer. Although genes for breast cancer has been identified, i.e. BRCA1 and BRCA2 among others, the attributable risk for these genes appears to be low. In fact it is unlikely that women with only one or even two family members affected with breast cancer are carriers of such genes. It is therefore of interest to clarify the role of a family history of breast cancer in risk prediction for non-carriers of the known susceptibility alleles as these women are likely to represent the majority of instances seen in clinic populations.

3. Towards a Comprehensive Model of Breast Cancer Risk

Parmigiani, Giovanni, Duke University

Address: Giovanni Parmigiani Box 90251, ISDS Duke University Durham, NC 27708-0251

Phone: 919-684-8064

Fax: 919-684-8594

Email: gp@stat.duke.edu

Abstract: Accurate breast cancer risk assessment is important for women who consider themselves at high risk for breast cancer and who seek counseling for decision-making purposes regarding prevention and for physicians interested in counseling women about their risks for breast cancer and their options. Also, it is important to improve accurate perceptions about risk among women

are not currently seeking counseling, such as first degree relative of breast cancer patients, who may over- or under-estimate their risks. This talk describes our progress to date in developing a comprehensive model for risk assessment in breast cancer. The model has two main goals: including recent advances in genetic susceptibility to breast cancer, and incorporating evidence from several data sets, including a cohort study. The talk will cover some technical aspects of our modeling effort as well as broader issues regarding the current and future role of statistical risk models in cancer prevention and counseling.

List of speakers who are nonmembers: None

29.4 asa.other.ass.05

The American College of Epidemiology: co-sponsored by Statistics in Epidemiology Section of the ASA

Session Slot: 10:30-12:20 Monday
Estimated Audience Size: 80 - 150
AudioVisual Request:

Session Title: Statistical Evaluation of Vaccine Effects

Vaccination is the most effective and cost-efficient way to control an infectious disease. The basic vaccine efficacy statistic has been around for over 80 years. However, the assessment of the properties of this statistic and the development of alternative measures of vaccine effects are relatively new research topics for biostatisticians and epidemiologists.

This session will focus on issues related to the statistical evaluation of vaccine effects, such as sample size and power calculations, errors in measurement of exposure status and the use of stochastic models in making policies regarding the design of vaccination programs.

Theme Session: no
Applied Session: yes

Session Organizer: **Haber, Michael** Emory University

Address: Department of Biostatistics Rollins School of Public Health Emory University Atlanta, GA 30322
Phone: 404-727-7698
Fax: 404-727-1370
Email: mhaber@sph.emory.edu

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 0 minutes First Speaker - 25 minutes Second Speaker - 25 minutes Third Speaker - 25 minutes Fourth Speaker - 25 minutes Floor Discusion - 10 minutes

Session Chair: **Haber, Michael** Emory University

Address: Department of Biostatistics Rollins School of Public Health Emory University Atlanta, GA 30322
Phone: 404-727-7698
Fax: 404-727-1370
Email: mhaber@sph.emory.edu

1. Preventing Epidemics in Heterogeneous Communities

Britton, Tom, Uppsala University

Address: Department of Mathematics Uppsala University Box 480, S-751 06 Uppsala, Sweden
Phone: 46-18-471-3222
Fax: 46-18-471-3201
Email: tom.britton@math.uu.se

Abstract: The main practical motivation for the study of epidemic models lies in the insight they provide about the control of infectious diseases. The present paper concerns estimation procedures for the basic reproduction number, R, in closed heterogeneous communities. R is important when designing vaccination policies in terms of how many and which individuals should be vaccinated. The estimation procedures are based on stochastic

models, thus allowing confidence regions, incorporating both individual heterogeneities, e.g. age and sex, and social heterogeneities such as households.

2. *Exact Power and Sample Sizes for Vaccine Efficacy Studies*

Chan, Ivan S. F., Merck Research Laboratories

Address: Merck Research Laboratories BL 3-2 West Point, PA 19486
Phone: 610-397-3391
Fax: 610-397-2931
Email: ivan_chan@merck.com
Bohidar, Norman R., Merck Research Laboratories

Abstract: In vaccine efficacy studies, the goal is to show that the vaccine reduces the incidence of the disease of interest when compared to placebo. Various methods of calculating sample size and power based on large sample approximation are available for these efficacy trials (Farrington and Manning 1990; Blackwelder 1993). In this paper we describe two procedures for calculating sample size and power based on exact distributions. In small studies where the disease incidence and the anticipated vaccine efficacy are both high, an unconditional exact procedure may be desirable because it guarantees the level of the test and loses little sensitivity. In large studies where the disease is rare, a Poisson approximation to the number of events is reasonable and an exact test is simple to construct conditional on the total number of events. One can use this exact procedure in study planning to guarantee the power, and the Poisson model is readily extended to handle different follow-up times between groups. We compare the power and type I error rate of these two exact methods to the method based on the normal approximation for varying disease incidence and sample size.

3. *Semiparametric Models for Mismeasured Exposure Information in Vaccine Trials*

Golm, Gregory T., Emory University

Address: Department of Biostatistics Rollins School of Public Health Emory University Atlanta, GA 30322

Phone: 404-727-7697

Fax: 404-727-1370

Email: golm@sph.emory.edu

Halloran, M. Elizabeth, Emory University

Longini, Ira M., Emory University

Abstract: Exposure to infection information is important for estimating vaccine efficacy, but is difficult to collect and inherently prone to missingness and mismeasurement. It is therefore not feasible to collect good exposure information on all participants in a large vaccine trial. We discuss study designs which collect detailed exposure information for only a small subset of trial participants, while collecting crude exposure information on all participants, and treat estimation of vaccine efficacy in the missing data/measurement errors framework. We demonstrate with the example of an HIV vaccine trial the improvements in bias and efficiency when combining different levels of exposure information to estimate vaccine efficacy for reducing both susceptibility and infectiousness. We compare the performance of recently developed semiparametric missing data methods of Pepe and Fleming (1994) and Robins, Hsieh and Newey (1995).

4. Statistical Evaluation of Preventive Vaccine Efficacy: Current Methods

Horhe, A. Dale, FDA/CBER

Address: FDA/CBER HFM-215 1401 Rockville Pike Rockville, MD 20852-1448

Phone: 301-827-6068

Fax: 301-827-3529

Email: HORNE@cber.fda.gov

Blackwelder, William C., NIH/NIAID

Abstract: Preventive vaccines are generally given to healthy humans to prevent disease or infection. Consequently, a relatively low risk/benefit ratio is required compared to many therapeutic agents. Thus, statistical evaluation depends primarily on estimation of efficacy, with a suitably high lower bound on the corresponding confidence interval. (A p-value from testing a null hypothesis of no difference from placebo is generally of little interest.)

Vaccine efficacy is routinely estimated as 1-R, where R may be a ratio of risks, incidence rates, or hazards in the vaccinated relative to control subjects. There are several methods for estimating R and for obtaining its confidence interval, which will be reviewed here.

Traditionally, preventive vaccine efficacy trials have not used an intention-to-treat (ITT) approach as the primary analysis. The reason for this practice is that many preventive vaccine trials give little cause for concern about bias due to dropouts and other sources of noncompliance, unlike most therapeutic trials. However, ITT is always an appropriate analysis to perform (perhaps as a co-primary analysis) and in some vaccine trials in which there may be considerable concern about bias, ITT may be the only appropriate analysis.

Sometimes it is not possible to evaluate vaccine efficacy directly via observation of who succumbs to the disease of interest. In such trials, which will be reviewed here, immunological surrogates of efficacy are used. Vaccine efficacy trials which use these surrogate endpoints are often problematic when no immunological correlate of protection is known. These trials are often designed and analyzed as equivalence trials.

List of speakers who are nonmembers: None

29.5 asa.other.ass.06

National Research Center for Statistics and the Environment: co-sponsored by Statistics and the Environment Section of the ASA

Session Slot: Monday, 2:00-3:50
Estimated Audience Size: 90-100
AudioVisual Request: overhead and slide projectors

Session Title: Influence of Statistical Considerations on Environmental Policy

Theme Session: yes
Applied Session: yes

Session Organizer: **Guttorp, Peter** University of Washington

Address: National Research Center for Statistics and the Enivronment Box 351720 University of Washington Seattle, WA 98195-1720
Phone: 206-616-9262
Fax: 206-616-9443
Email: peter@stat.washington.edu

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 5 minutes First Speaker - 30 minutes Second Speaker - 30 minutes Third Speaker - 30 minutes Floor Discusion - 15 minutes

Session Chair: **Cox, Larry** U.S. Environmental Protection Agency

Address: U.S. Environmental Protection Agency National Exposure Research Laboratory(MD-75) Research Triangle Park, NC 27711
Phone: 919-541-2648
Fax: 919-541-7588
Email: cox.larry@epamail.epa.gov

1. Setting Environmental Standards: A Statistical Evaluation of the US Ozone Standard

Thompson, Mary Lou, University of Washington

Address: Dept. of Biostatistics University of Washington H655G Health Sciences Box 357232 Seattle, WA 98195-7232
Phone: (206) 616-2723

Fax: (206) 685-7419
Email: mlt@biostat.washington.edu
Guttorp, Peter, University of Washington
Sampson, Paul D., University of Washington
Cox, Larry, EPA

Abstract: The US Environmental Protection Agency recently revised the regulations in the clean air act regarding air pollution, setting new standards for ozone and for particulate matter. we discuss a general approach to the setting of standards from the perspective of statisticians and other environmental scientists, and illustrate some of the issues with ozone data from California's central valley.

2. *Statistics and Mandated Science*

Van Belle, Gerald, University of Washington

Address: Department of Environmental Health Box 357234 University of Washington Seattle, WA 98195-7234
Phone: 206-543-6991
Fax:
Email: vanbelle@u.washington.edu

Abstract: Mandated science operates at the intersection of science and policy, both which are embedded in values. mandated science defines, directs and uses the results of science in establishing and evaluating policy. statisticians and statistical methods play a key role in mandated science. this paper illustrates some key characteristics of statistical methods used in mandated science. these results are then applied to three examples of mandated science: lead in soils, air pollution and clean up of the nations nuclear weapons complex. the statistical issues in mandated science include problems of measurement, exposure, association, the use of large data bases not specifically established for scientific purposes, and linkages of data bases using statistical rather than mechanistic models.

3. Statistics in Environmental Decision Making: They Really Do Count!

Nussbaum, Barry D., EPA

Address: Center for Environmental Information and Statistics Mail Code 2163 U.S. Environmental Protection Agency 401 M Street, SW Washington, D.C. 20460

Phone: (202) 260-1493

Fax: 202-260-4968

Email: NUSSBAUM.BARRY@EPAMAIL.EPA.GOV

Abstract: Despite the cynicism, statistics are used in the decisions made on environmental policy, regulations, and enforcement actions. However, it is frequently the effectiveness of the presentation of the statistics, rather than the raw statistics that make the difference. several general guidelines are explored; and then three specific examples in which the author was personally involved are discussed. one ended up in a regulation, one in court, and one on the desk of the President of the United States.

List of speakers who are nonmembers: None

29.6 asa.other.ass.07

Federal Science Agencies

Session Slot: 8:30-10:20 Tuesday
Estimated Audience Size: 60 –120
AudioVisual Request:

Panel Session Title: Federal Funding for Statistical Research

Theme Session: No
Applied Session: No

Session Organizer: **Gentle, James** National Science Foundation

Address: Division of Mathematical Sciences
National Science Foundation
4201 Wilson Blvd., Room 1025
Arlington, VA 22230
Phone: 703-306-1883
Fax: 703-306-0555
Email: jgentle@nsf.gov

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 5 minutes First Panelist - 15 minutes Second Panelist - 15 minutes Third Panelist - 15 minutes Fourth Panelist - 15 minutes Fifth Panelist - 15 minutes Sixth Panelist - 15 minutes Floor Discusion - 15 minutes (or 5 or 15)

Session Chair: **Gentle, James** National Science Foundation

Address: Division of Mathematical Sciences
National Science Foundation
4201 Wilson Blvd., Room 1025
Arlington, VA 22230
Phone: 703-306-1883
Fax: 703-306-0555
Email: jgentle@nsf.gov
Launer, Robert L. U. S. Army Research Office

Address: Mathematical and Computer Sciences Division
P.O. Box 12211
Research Triangle Park NC 27709-2211
Phone: (919) 549-4309
Fax: (919) 549-4354
Email: launer@aro-emh1.army.mil
Cox, Larry US EPA

Address: US EPA, NERL (MD-75)
Research Triangle Park, NC 27711
Phone: 919 541-2648
Fax: 919 541-7588

Email: cox.larry@epamail.epa.gov
Albert, Jeffrey M. NIAID

Address: NIH/NIAID/DAIDS/BRB
6003 Executive Blvd., MSC 7620
Rm 2A26B
Bethesda, MD 20892
Phone: (301) 435-3752
Fax: (301) 480-5703
Email: ja24o@nih.gov
Maar, Jim National Security Agency

Address: Dr. James R. Maar, External Research Director
Mathematical Research Division, R51
National Security Agency
9800 Savage Road, STE 6515
Ft. Meade, MD 20755-6515
Phone: 301-688-0105
Fax: 301-688-0689
Email: jrmaar@afterlife.ncsc.mil
Crank, Keith National Science Foundation

Address: Division of Mathematical Sciences
National Science Foundation
4201 Wilson Blvd., Room 1025
Arlington, VA 22230
Phone: 703-306-1885
Fax: 703-306-0555
Email: kcrank@nsf.gov
Poston, Wendy Office of Naval Research

Address: 800 North Quincy Street
Code 311, BCT-1
Arlington, VA 22217
Phone: 703-696-4320
Fax: 703-696-2611
Email: postonw@onr.navy.mil
List of speakers who are nonmembers: None

Chapter 30

General Methodology (4)

30.1 asa.pres.invited.01

Sponsoring Section/Society: ASA

Session Slot: 4:00- 5:50 Monday

Session Title: Invited Paper Session Title Here

Session Organizer: **Moore, David S.** Purdue University

Session Chair: **Moore, David S.** Purdue University

1. TBA
Moynihan, Patrick Daniel, Honorable Senator from New York

Discussant: **TBN**

Discussant: **TBN**

30.2 asa.deming.01

Sponsoring Section/Society: ASA

Session Slot: 4:00- 5:50 Tuesday
Estimated Audience Size: 500

AudioVisual Request: Video projector for his Macintosh Laptop

Session Title: Deming Lecture

Myron Tribus, PE, is a consulting engineer, specializing in Quality Management and is a Director of Exergy, a company specializing in the design of advanced high efficiency power production systems. He retired from MIT in 1986 after 11 1/2 years as Director of the Center for Advanced Engineering Study. Before coming to MIT he was a Senior Vice President for Research and Engineering for the Xerox Corporation, where he was in charge of Research, Development and Engineering for the entire line of Xerox Copiers, Duplicators and Telecopiers. Before that he served for two years as Assistant Secretary ofor Science and Technology in the U. S. Department of Commerce. For eight years he was Dean of the Thayer School of Engineering at Dartmouth College where he introduced the Unified Engineering Curriculum and lead the faculty in developing a curriculum based on engineering design and entrepreneurship. Dr. Tribus served for 16 years on the faculty of the College of Engineering and UCLA and two years on the Faculty of the University of Michigan. He has worked as a design engineer for the Jet Engine Development of the General Electric Company. In 1958 he hosted the television show “Threshold” for CBS in Los Angeles.

Dr. Tribus has published over 100 papers on topics ranging from academic subjects, such as heat transfer, fluid mechanics, probability theory, statistical inference and thermodynamics to applied topics such as sea water demineralization, aircraft heating, aircraft ice prevention and the design of engineering curricula. He has published two books, “Thermostatics and Thermodynamics” (which provided the first textbook which bases the laws of thermodynamics on information theory instead of the classical arguments) and “Rational Descriptions, Decisions and Designs” (which introduces Bayesian Decisions methods into the engineering design process). Dr. Tribus was active with Irving Langmuir in the pioneering days of cloud seeding and has published analyses of the role of statistics in weather modification.

Dr. Tribus has also published and lectured extensively on topics of social interest such as the position of engineers in politics, the decline of US competitiveness in world trade, the role of decision theory in political decision making and the role of technology in society.

In recent years dr. Tribus has become known through his writings on Dr. Deming’s philosophy of management. He is one of the founders of the

American Quality and Productivity Institute which is devoted to teaching and promoting the fundamentals of quality management. The AQPI merged its efforts with the Transformation of American Industry project of Jackson Community College to form the Community Quality Council to foster the growth and development of community quality centers across the USA. This activity has recently been merged with the activities of the Association for Quality and Participation, which will continue the support of community quality centers.

Dr. Tribus is currently a partner in Exergy, an enterprise devoted to the commercialization of a new power cycle for utilities to use in electric power generation, the Kalina Cycle, which promises significant improvements in efficiency and economy, using off-the-shelf hardware and conventional levels of pressure and temperature. The new technology is currently being commercialized. Worldwide rights to the "bottoming cycle" have been acquired by the General Electric Company. The patented design for geothermal applications increases the output of a geothermal bore by as much as 40%. In a recent report the U.S. Department of Energy reported that this technology doubles the amount of land in the USA which may be exploited for geothermal electrical energy production. This design has been licensed to Ansaldo, an Italian company which holds the largest share of the international geothermal market. The technology for coal and other solids, which promises a reduction in coal use by 20%, has been licensed by Ebara, a Japanese company. A commercial installation based on the incineration of industrial waste is currently in the design stages.

Dr. Tribus is a member of the National Academy of Engineering. He has served on several boards of directors and been a consultant to many companies and governmental bodies. He has received five awards for technical papers and public service and two honorary doctorates. He received the BS in Chemistry from the University of California, Berkeley in 1942 and the PhD in Engineering from UCLA in 1949.

Dr. Tribus is married and has two grown daughters.

Session Organizer: **Deming Lectureship Committee**

Session Chair: **Starbuck, Robert R.** Wyeth-Ayerst Research

Address: Clinical Research & Development Wyeth-Ayerst Research PO Box 42528 Philadelphia, PA

Phone: 610-341-2070
Fax: 610-341-2092
Email: starbur@war.wyeth.com

1. The Contributions of W. Edwards Deming to the Improvement of Education

Tribus, Myron, Exergy, Inc.

Address: 350 Britto Terrace Fremont, CA 94539
Phone: 510-651-3641
Fax: 510-656-9875
Email: 104055.2663@compuserve.com

Abstract: *W. Edwards Deming's teachings about management and about education have inspired several teachers to change the way they approach their jobs. The transformation began with one school in Alaska in the 1980's and has since spread world-wide with schools as far away as Tasmania and Latin America working at the transformation. The trials and tribulations along the way resemble those encountered in industry. The greatest strides have been made at the lower grades; universities have thus far, with minor exceptions, given only lip service.*

When Deming's ideas about management are combined with Reuven Feuerstein's practices in cognitive development, the promise of a wholly new and vastly improved educational experience for future generations is in sight. To what degree this vision can be attained will depend upon the continued spread of the ideas of these two men. In the words of Dr. Deming, success will require "a critical mass of people who understand and work consistently."

30.3 asa.pres.address.01

Sponsoring Section/Society: ASA

Session Slot: 8:00-10:00 Tuesday
Estimated Audience Size: xx-xxx
AudioVisual Request: xxx

Session Title: ASA Presidential Address

Theme Session: Yes
Applied Session: Yes/No

Session Organizer: **Moore, David S.** Purdue University

Address: Dept Of Statistics Purdue University W Lafayette, IN 47907-1399
Phone: 765-494-6050
Fax: 765-494-0558
Email: dsm@stat.purdue.edu

Session Timing: 110 minutes total (Sorry about format):

Opening Remarks by Chair - 0 First Speaker - 30-60 minutes
Floor Discussion - 0

Session Chair: **Kettenring, Jon R.** BellCore

Address: Bellcore 445 South St Morristown, NJ 07960-6438
Phone: 201-829-4398
Fax: 201-829-4391
Email: jon@bellcore.com

1. ***Statistics Among the Liberal Arts***
Moore, David S., Purdue University

Address: Dept Of Statistics Purdue University W Lafayette, IN 47907-1399
Phone: 765-494-6050
Fax: 765-494-0558
Email: dsm@stat.purdue.edu

Abstract:

List of speakers who are nonmembers: None

30.4 asa.fisher.01

Sponsoring Section/Society: COPPS

Session Slot: 4:00- 5:50 Wednesday
Estimated Audience Size: xx-xxx
AudioVisual Request: xxx

Session Title: Fisher Lecture

Theme Session: Yes/No
Applied Session: Yes/No

Session Organizer: **TBN**

Address:
Phone:
Fax:
Email:

Session Timing: 110 minutes total (Sorry about format):

110 minutes total...please allocate Opening Remarks by Chair - 5 or 10 minutes Speaker - 30 or 45 minutes Discussant - 10 minutes (or none) Discussant - 10 minutes (or none)

Session Chair: **TBN**

Address:
Phone:
Fax:
Email:

1. Logicist Statistics: Modeling and Inference
Dempster, Arthur P., Harvard University

Address: Dept of Stat, Science Center Harvard Univ Cambridge, MA 02138-2901
Phone: 617-495-5601

Fax:
Email: dempster@stat.harvard.edu

Abstract: *An alternative view of the practice of statistics is presented in this lecture, stressing roles for both formal and informal reasoning about specific situations. Whereas most expositions of our subject put a dominating focus on procedures for data gathering, data analysis, and inference that emphasizes their mathematical, computational, and reporting aspects, the approach called here logicist stresses that procedures are also packaged forms of reasoning, (i.e., "logic" in a broad sense), and insists that for applications such logic is the sine qua non of statistical analysis. Formal statistical reasoning depends on formal models that may be (i) "empirical" based on data analysis, or (ii) "stochastic" representing repetitive phenomena such as sampling, measurement, or temporal-spatial variation, or (iii) "predictive" aimed at probabilistic reasoning about unknowns. These types are strongly overlapping. The basic inferential mechanisms are postdictive interpretation of model-dependent probabilities associated with observed outcomes, as with significance tests or likelihood ratio interpretations, and prospective interpretation of probabilities of unknown factual assertions computed from fusion of data and other information, as with interpretation of Bayesian or belief function posteriors. Procedures for model assessment from data, including model testing, selection, and sensitivity are important, but are only half the story of model construction. Independence judgments, and prior understanding and knowledge, are also fundamental. The logicist attitude to probability and to statistical science originates in the scientific outlook of R. A. Fisher, and is sharply different from narrower "frequentist" or "Bayesian" schools which neither individually nor jointly support enough of inferential logic to carry us forward into a future of burgeoning technological complexities.*

Discussant: **TBN**

Address:

Phone:
Fax:
Email:

Discussant: **TBN**

Address:
Phone:
Fax:
Email:
List of speakers who are nonmembers: None

Chapter 31

ASA Competition (10 – completed 11/25/97)

Chapter 32

Invited Posters (Number to be Determined)

Program Chair: **Scott, David W.**

32.1 iposter.01

INVITED POSTER – ASA Stat Ed Section

Session Slot: Sunday, 4:30 - 7:00
AudioVisual Request: None

Session Organizer: **Moreno, Jerry**
Invited Poster Title: Linking Nursing Research Datasets to a Nursing Statistics Course through a Web Archive

Short, Thomas H., Villanova University

Address: Department of Mathematical Sciences Villanova University 800 Lancaster Ave. Villanova, PA 19085-1699
Phone: 610-519-6961
Fax: 610-519-6928
Email: short@monet.vill.edu
Fetter, Marilyn S., Villanova University

Abstract: Statistical consulting projects with researchers in the College of Nursing at Villanova have provided a pool of datasets that can be shared with students in a required statistics course for undergraduate Nursing majors.

Since the datasets tend to be large and unwieldly, we have installed them in a Web archive that allows students to select subsets of variables and observations that can be analyzed using tools learned in a first statistics course.

We will summarize the pedagogical issues involved in the use of the data archive, and review experiences we have had in our courses that use the archive.

32.2 iposter.02

Sponsoring Section/Society: Health Policy Statistics Section

Session Slot: Sunday, 4:30 - 7:00
AudioVisual Request: None
Invited Poster Title: The Clustered Encouragement Design

Frangakis, Constantine, Harvard University

Address: Harvard University, Department of Statistics, Science Center, 1 Oxford St., Cambridge MA 02138
Phone: (617) 495 4888
Fax: (617) 496 8057
Email: frangaki@hustat.harvard.edu
Rubin, Donald B., Harvard University
Zhou, Xiao-Hua (Andrew), Indiana University

Abstract: Comparative study designs where the encouragement, rather than direct receipt of treatment, is randomly assigned have attracted attention in settings where traditional randomization of receipt is not ethical or not feasible (e.g., Hirano, Imbens, Rubin, and Zhou, 1997). Combining this encouragement design with the practical constraint of clustered randomization (e.g., at *the physician level), generates a relatively simple design with potential to evaluate the effectiveness of health-care-related treatment options. Unfortunately, the design does not prevent indirect effects, of no policy interest, from being present, which renders standard intention-to-treat or exclusion-restriction-based instrumental variables analyses (e.g., Angrist, Imbens, and Rubin, 1996) generally inappropriate or inapplicable. In this work, we develop an analytic method to address this situation by bringing together and extending ideas from recent work on encouragement designs (Hirano et al., 1997), while allowing for interference between units (Rubin, 1978) to accommodate the clustered randomization. We show that the resulting procedure has key theoretical advantages over existing ones and that, despite its multilevel structure, still it can be relatively easily implemented using current simulation techniques. We apply our methods to a study on the effectiveness of physician-patient discussions of Advanced Directive forms, and show how the theoretical results are reproducible in real studies, thereby supporting the clustered encouragement design as a practical tool.*

Theme Session: Yes
Applied Session: Yes

32.3 iposter.03

Sponsoring Section/Society: Health Policy Statistics Section

Session Slot: Sunday, 4:30 - 7:00
AudioVisual Request: None
Invited Poster Title: Who Doesn't Get a Pap Test?

Gentleman, Jane F., Statistics Canada

Address: Social Survey Methods Division, Statistics Canada R.H. Coats Bldg, 15th floor, Tunney's Pasture Ottawa, Ontario, Canada K1A 0T6
Phone: (613) 951-8553
Fax: (613) 951-0653

Email: gentjan@statcan.ca
Lee, Judy, Statistica Canada
Parsons, Greg F., Statistics Canada

Abstract: This study examines the characteristics associated with getting or not getting a Pap test. Knowing who does not get a Pap test can help health policy makers target their resources and evaluate the effectiveness of their programs for encouraging compliance with cervical cancer screening guidelines. Canadian guidelines recommend that most women get Pap smears every three years after two satisfactory smears one year apart. The data analyzed were from Statistics Canada's 1994/95 National Population Health Survey. Multivariate logistic regressions were run to determine the odds of women ever having had a Pap test and, among women who had, the odds of having done so in the previous one and three years. The results show that the odds of ever having had a Pap test increase with age up to age group 30-39, after which they decrease. The odds for women aged 70 and over are only 70% of those for women aged 18-19. It might have been expected that the oldest women would be the most likely to have ever had a Pap test, since they have had the longest opportunity to have done so. The lower Pap test rates for older women are evidently due to differing attitudes toward and opportunities for Pap tests. Other characteristics of women with low odds of having had a Pap test are: being single (never married), being a resident of the province of Quebec, having immigrated to Canada, having less than a secondary level of education, not being in the highest household income group, and having had relatively few sex partners in the last year. Overall, 26% of women in the targeted age group (18-69) were not adhering to current screening guidelines because they had either not ever had a Pap test or they had not had one in the last three years.

Theme Session: Yes
Applied Session: Yes

32.4 iposter.04

Health Policy: Co-Sponsored by (TBA)

Session Slot: Sunday, 4:30 - 7:00
AudioVisual Request: None

Invited Poster Title: Issues in Case-Mix Adjustment of Measures of the Quality of Health Plans

Zaslavsky, Alan,

Address:
Phone:
Fax:
Email:

Abstract: In a health care environment increasingly dominated by managed care, measurements of the quality of care provided by health plans are important tools for consumer and purchaser choice. Two important quality measurement tools are the CAHPS survey of health plan consumer satisfaction, and the HEDIS data set, which compiles rates at which essential preventative services are provided to health plan consumers. Because the health plans may select for patients with different characteristics, comparisons between plans should be adjusted to allow for the effects of case mix on plan ratings. Results for case-mix adjustment of CAHPS and HEDIS are reported. Some issues of methodology and interpretation are considered, including the use of ecological variables for case mix, the implications of nonparallelism of the case-mix regressions, and presentation of the magnitude of the case-mix effects.

Theme Session: No
Applied Session: No

32.5 iposter.05

Business & Economics

Session Slot: Sunday, 4:30 - 7:00
AudioVisual Request: None
Invited Poster Title: TBA

Bruce, Andrew,

Address:
Phone:
Fax:
Email:
,

Abstract:

Theme Session: Yes/No
Applied Session: Yes/No

32.6 iposter.06

ASA Epidemiology

Session Slot: Sunday, 4:30 - 7:00
AudioVisual Request: None
Invited Poster Title: TBA

Hoffmann, Ray,

Address:
Phone:
Fax:
Email:

Abstract:

Theme Session: Yes/No
Applied Session: Yes/No

32.7 iposter.07

Section on Statistical Consulting

Session Slot: Sunday, 4:30 - 7:00
AudioVisual Request: Television monitor, VCR
Invited Poster Title: Statistics in Consumer Reports

Burn, David, Consumers Union

Address: 101 Truman Avenue, Yonkers, NY 10703
Phone: 914-378-2306
Fax: 914-378-2908
Email: burnda@consumer.org
Romm, Marty, Consumers Union

Abstract: If you're like millions of Americans, whether you want to buy a toaster or a car, you turn to the pages of Consumer Reports. You may simply want to know what the best rated product is, or you may look for more detailed information on product attributes and how they relate to performance. And you trust in the guidance you find because you know Consumers Union, publisher of Consumer Reports, is a nonprofit, independent consumer information organization with a long history of unbiased and rigorous consumer product testing. In this presentation, we take a 'behind-the-scenes' glimpse of product testing at Consumers Union, and show the vital role statisticians play in crunching much more than just numbers.

32.8 iposter.08

Biostatistics Section

Session Slot: Sunday, 4:30 - 7:00
AudioVisual Request: None
Invited Poster Title: The Concordance Correlation Curve

Lombard, Carl J., ASA

Address: Centre for Epidemiological Research in South Africa, Medical Research Council, P O Box 19070, Tygerberg 7505, Cape Town, South Africa
Phone: 27-021-9380328
Fax: 27-021-9380310
Email: clombard@eagle.mrc.ac.za

Abstract: The concordance correlation coefficient was developed as an index of reproducibility in studies or as an index to assess agreement in methods comparison. The index consists of an accuracy and precision component and the latter is measured by the ordinary Pearson correlation coefficient. The Pearson correlation coefficient has been generalized to the case of nonlinear hetroscedastic regression in the form of a correlation curve. This correlation curve provides a local measure of association over the range of covariate values. The concordance correlation curve is based on this approach and provides a measure of local concordance. The index will be illustrated with the comparison of two methods for the measurement of sperm morphology.

Theme Session: No
Applied Session: Yes

32.9 iposter.09

Statistical Computing

Session Slot: Sunday, 4:30 - 7:00
AudioVisual Request: An LCD projection device and an internet connection if possible
Invited Poster Title: Computing with WebStat

West, Webster, University of South Carolina

Address: Department of Statistics University of South Carolina Columbia SC 29208
Phone:
Fax: (803) 777-3792
Email: (803) 777 -4048
Ogden, Todd, University of South Carolina

Abstract: WebStat, a statistical computing environment for the World Wide Web was originally developed so that students might have access to basic statistical procedures on the Internet. The developers of WebStat have continued to add new features and procedures to the package in order to make it more functional for a larger number of potential users. These newest capabilities will be on display as the latest version of WebStat will be made available for interested users to try out for themselves. In addition to the software demonstration, plans for the future development of WebStat will be discussed.

32.10 iposter.10

To be determined...

Chapter 33

Invited Classroom Day Presentations

Program Chair: **Scott, David W.**

33.1 iclass.01

CLASSROOM DAY

Session Slot: Monday, 8:30 - 9:25
AudioVisual Request: None

Session Organizer: **Scott, David W.**

Session Title: Classroom Day

1. Statistical Process Control: Investigating Sources of Variation

Sylwester, David, University of Tennessee at Knoxville

Address: Department of Statistics University of Tennessee, Knoxville, TN 37996-0532
Phone: 423-974-8326
Fax: 423-974-2490
Email: dsylwest@utk.edu

Abstract: *This overview of statistical process control (SPC) is part of Classroom Day. It is intended for statisticians who work in business and industry as well as those who teach MBA students.*

SPC is usually viewed as concentrating on control charts with (undue) emphasis on points outside of control limits. A broader, more useful, view is to consider it as the study of sources of variation. In this tutorial I will introduce the basic statistics tools of SPC and then show examples in which analysis is pointed at investigating and exploiting sources of variation. My experiences in teaching MBA students about statistics through SPC will be briefly mentioned.

Keywords: Statistical process control, investigating variation

AudioVisual Request: None

33.2 iclass.02

CLASSROOM DAY

Session Slot: Monday, 9:30 - 10:25
AudioVisual Request: None

Session Organizer: **Scott, David W.**

Session Title: Classroom Day

1. Graphical Methods for Design of Experiments
Barton, Russell R., The Pennsylvania State University

Address: Department of Industrial and Manufacturing Engineering The Pennsylvania State University 207 Hammond Building University Park, PA 16802
Phone: 814-863-7289
Fax: 814-863-4745
Email: barton@simplex.psu.edu

Abstract: Experiment design involves defining experimental goals, identifying and classifying key variables, postulating a probability model, selecting experimental runs, and verifying the ability of the design to fit the proposed model. This talk presents an overview of graphical methods for each of these activities which are easy to use and easy to teach. As a bonus, the graphical framework used to display factorial and fractional-factorial designs with up to ten factors can be used to present experimental results in a model-free context which can sometimes provide more insight than a fitted regression model. This talk is part of the classroom day program.

KeyWords: graphical displays, goal hierarchy plots, fractional factorial designs.

AudioVisual Request: None

33.3 iclass.03

CLASSROOM DAY

Session Slot: Monday, 10:30 - 11:25
AudioVisual Request: None

Session Organizer: **Scott, David W.**

Session Title: Classroom Day

1. ***Statistics and Wall Street: Show Me the Money***
   Welsch, Roy E., Sloan School of Management, MIT

Address: Massachusetts Inst of Tech 50 Memorial Drive, E53-383 Cambridge, MA 02139-4307
Phone: 617-253-6601
Fax: 617-258-7579
Email: rwelsch@mit.edu

Abstract: This is a Classroom Day lecture and will survey some of the statistical issues that arise in modern finance along with enough finance to understand the problems addressed. Topics: measuring risk and return (are variances and means enough?), value-at-risk (should we forget three sigma?), diversification (what's so great about the covariance matrix?), portfolio optimization and sensitivity analysis (diagnostics), "beta" and the capital asset pricing model (why is linear least-squares always best?), and a few words about options. Some possible research topics will be mentioned along with ideas about what a statistician might do on Wall Street (other than get higher pay). Any investment advice will be strictly random.

Key Words: finance, risk, value-at-risk, beta, CAPM, portfolio optimization

AudioVisual Request: None

33.4 iclass.04

CLASSROOM DAY

Session Slot: Monday, 11:30 - 12:25
AudioVisual Request: None

Session Organizer: **Scott, David W.**

Session Title: Classroom Day

1. Statistics and the Search For Disease Genes

Weir, Bruce S., North Carolina State University

Address: Department of Statistics North Carolina State University Raleigh NC 27695-8203
Phone: 919-515-3574
Fax: 919-515-7315
Email: weir@stat.ncsu.edu

Abstract: A review of current statistical methods for locating human disease genes will be given as part of the classroom day program. One of the outcomes of the human genome project has been the development of large numbers of easily-scored "genetic markers." If markers at known locations on the human genome are found to have an association with disease status in a population, this may be because the disease gene is close to the marker location. It may also be a consequence of the evolutionary history of the population. Among the many statistical issues that need to be addressed is that of correcting for multiple tests when as many as 60,000 markers are examined. The variation in levels of association due to evolution must also be accommodated.

Key words: Genome project, human disease, genetics

AudioVisual Request: None

33.5 iclass.05

CLASSROOM DAY

Session Slot: Monday, 2:00 - 2:55
AudioVisual Request: None

Session Organizer: **Scott, David W.**

Session Title: Classroom Day

1. Quotient Spaces in Statistical Models

McCullagh, Peter, University of Chicago

Address: Department of Statistics University of Chicago 5734 University Ave Chicago, IL 60637
Phone: 773-702-8340
Fax: 773-702-9810
Email: pmcc@roxbee.uchicago.edu

Abstract: A statistical variable, or variate for short, is a function, usually real-valued, on the statistical units Ω. Each variate is thus an element in the vector space $\mathcal{V} = \mathcal{R}^{\Omega}$. Many statistical models, including all linear models and generalized linear models, are specified in terms of subspaces, usually in the form $\mu \in \mathcal{X}$, $\log \mu \in \mathcal{X}$, or $g(\mu) \in \mathcal{X}$, where $\mathcal{X}$ is the span of the model matrix X. Quotient spaces are rarely considered explicitly in statistical work, but it will be argued in this talk that this oversight is a mistake. Quotient spaces do arise naturally in various settings of which the following are typical examples.

In the case of the linear model $E(Y) \in \mathcal{X}$, *the residual* $Y + \mathcal{X}$ *lies in the quotient space* $\mathcal{V}/\mathcal{X}$. *An elementary calculation shows that* $Y + \mathcal{X}$ *is normally distributed in the quotient space, whatever this may mean, and the likelihood function based on* $Y + \mathcal{X}$ *is the REML likelihood.*

In testing a composite null hypothesis H_0 *against an alternative* $H_A \supset H_0$, *if both both hypotheses correspond to vector subspaces, the standard test statistics are based on the quotient space projection on to* H_A/H_0.

Interaction between two factors A, B, *is the quotient-space projection* $Y + (A+B)$ *of the response* Y *on to the space* $\mathcal{V}/(A+B)$, *i.e.* Y *modulo additivity. Regardless of whether the design is balanced, the interaction sum of squares is* $\|Y + (A+B)\|^2$, *or more generally* $\|PY + (A+B)\|^2$, *where* P *is the orthogonal projection on to* $A.B$.

In the case of multinomial response models, the probability vector $(\pi_1, \ldots, \pi_k)$ *is a point in the probability simplex in* $\mathcal{R}^k$. *The usual link transformation* $\eta_j = \log \pi_j$ *has inverse*

$$\pi_r = \exp(\eta_r) / \sum \exp(\eta_j)$$

so that η_j *and* $\eta_j + c$ *are equivalent points. In other words,* η *is a point in* $\mathcal{R}^k/1$, *and the logistic link is 1–1 from the interior of the simplex onto* $\mathcal{R}^k/1$.

Bayes's theorem is quotient-space vector addition.

For incomplete data, if the relation between the incomplete response and the complete response is linear, the incomplete response is most naturally viewed as a point in a certain quotient

space. Although more restrictive than the EM algorithm, this formulation leads to a simple theory of estimation using estimating functions.

The lecture will be in the form of an expository review session, beginning with the definition of a quotient space, continuing with inner products norms and Lebesgue measure on quotient spaces, and deriving the normal distribution on $\mathcal{V}/\mathcal{X}$. Details of this and other examples listed above can be found at http://galton.uchicago.edu/~pmcc/quotient.ps

KeyWords: interaction, analysis of variance, multinomial response model, REML

AudioVisual Request: None

33.6 iclass.06

CLASSROOM DAY

Session Slot: Monday, 3:00 - 3:55
AudioVisual Request: None

Session Organizer: **Scott, David W.**

Session Title: Classroom Day

1. Nonparametric Curve Estimation, Who Needs It?
Marron, J. S., University of North Carolina

Address: Department of Statistics University of North Carolina Chapel Hill, NC 27599-3260
Phone: 919-962-2188
Fax: 919-962-1279
Email: marron@stat.unc.edu

Abstract: This presentation will be part of the classroom day. This will be an introduction (with a stress on intuition, graphics and real data examples) to the basic ideas, and "essential folklore" *of smoothing, i.e. nonparametric curve estimation. Some topics in smoothing that have been recently debated in a heated and rather public way will be reviewed, from several points of view. Some very new ideas, such as scale space theory, that have the potential to seriously impact the way statisticians think about smoothing will also be discussed.*

KeyWords: nonparametric curve estimation, kernel smoothing

AudioVisual Request: None

Chapter 34

College Bowl (2)

34.1 college.bowl.01

Sponsoring Section/Society: Mu Sigma Rho and ASA Section on Statistical Education

Session Slot: Tuesday, 10:30 -12:30
Estimated Audience Size: 200
AudioVisual Request: At least three table mikes, preferably five. Three tables on stage, seating 4 each. Also, three large poster-hanging boards (similar to those used in Poster sessions).

Panel Session Title: College Bowl Quarterfinals

Four competitions of approximately 20 minutes' length, each involving two teams representing graduate Statistics programs (see article in Winter,1997 issue of STATS, p.17, for more details; or: http://www.stat.sc.edu/msrnatl.html.

Theme Session: No
Applied Session: No

Session Organizer: **Payton, Mark** Oklahoma State University

Address: Dept. of Statistics, MS 301 Oklahoma State University Stillwater, OK 74078-1056

Phone: (405)744-5684
Fax: (405)744-7074
Email: mpayton@okway.okstate.edu

Session Organizer: **Albert, Jim** Bowling Green State University

Address: Department of Mathematics and Statistics Bowling Green State University Bowling Green, OH 43403
Phone: 419-372-7456
Fax: 419-372-6092
Email: albert@math.bgsu.edu

Session Timing: 110 minutes total (Sorry about format):

Unique

Session Chair: **Albert, Jim** Bowling Green State University

Address: Department of Mathematics and Statistics Bowling Green State University Bowling Green, OH 43403
Phone: 419-372-7456
Fax: 419-372-6092
Email: albert@math.bgsu.edu

Moderator: **Woodward, Wayne** Southern Methodist University

Judge: **Stephenson, Bob** Iowa State University

Scorekeeper: **Gerow, Ken** University of Wyoming

Buzzermeister: **Edwards, Don** University of South Carolina
List of speakers who are nonmembers: None

34.2 college.bowl.02

Sponsoring Section/Society: Mu Sigma Rho and ASA Section on Statistical Education

Session Slot: Wednesday, 10:30 -12:30
Estimated Audience Size: 200
AudioVisual Request: At least three table mikes, preferably five. Three tables on stage, seating 4 each. Also, three large poster-hanging boards (similar to those used in Poster sessions).

Panel Session Title: College Bowl Semifinals and Finals

Four competitions of approximately 20 minutes' length, each involving two teams representing graduate Statistics programs (see article in Winter,1997 issue of STATS, p.17, for more details; or: http://www.stat.sc.edu/msrnatl.html.

Theme Session: No
Applied Session: No

Session Organizer: **Payton, Mark** Oklahoma State University

Address: Dept. of Statistics, MS 301 Oklahoma State University Stillwater, OK 74078-1056
Phone: (405)744-5684
Fax: (405)744-3533
Email: mpayton@okway.okstate.edu

Session Organizer: **Albert, Jim** Bowling Green State University

Address: Department of Mathematics and Statistics Bowling Green State University Bowling Green, OH 43403
Phone: 419-372-7456
Fax: 419-372-6092
Email: albert@math.bgsu.edu

Session Timing: 110 minutes total (Sorry about format):

Unique

Session Chair: **Albert, Jim** Bowling Green State University

Address: Department of Mathematics and Statistics Bowling Green State University Bowling Green, OH 43403
Phone: 419-372-7456
Fax: 419-372-6092
Email: albert@math.bgsu.edu

Moderator: **Boyer, John** Kansas State University

Judge: **McCulloch, Chuck** Cornell University

Scorekeeper: **Lancaster, Carol** Medical University of South Carolina

Buzzermeister: **Edwards, Don** University of South Carolina
List of speakers who are nonmembers: None