SACRAMENTO
STATISTICAL
ASSOCIATION

A local chapter of the American Statistical Association

PROGRAM

Institute on Research and Statistics

Friday, April 3, 2009
California State University, Sacramento

http://www.amstat.org/chapters/sacramento/

President: Rahman Azari, 530-752-7709; Vice-President: Shannon Conroy, 916-208-3161; Secretary: Raphael Diaz, 916-278-6221; Treasurer: Charles Chan, 552-9694; 2009 ASA Representative and Past President: Gloria J. Robertson, 916-654-1837 Councilors: Tadese Alemu, 916-322-4086; Rose Cendak, 530-757-7492, Thomas Stopka 916-449-5828; Lenard Swope, 707-463-7972; Farzaneh Tabnak, 530-753-0378

ALL EVENTS WILL TAKE PLACE AT THE CALIFORNIA STATE UNIVERSITY UNION

Registration and Continental Breakfast	Redwood Room	8:00- 9:00

Morning Plenary Sessions	Redwood Room	9:00-10:50

9:00-9:10

Welcoming Remarks
Rahman Azari, President, Sacramento Statistical Association
Jill M. Trainer, Dean, School of Natural Sciences and Mathematics, CSUS

9:10-10:55

Featured Speaker
Howard Roth, Chief Economist, California Department of Finance
2009: A Tough Year for the California Economy

10:55-10:40

Featured Speaker
Tim Carpenter, Professor, Department of Medicine & Epidemiology, and
Co-Director, Center for Animal Disease Modeling and Surveillance, UC Davis
To be announced

10:40-10:50

Featured Speaker
Wolfgang Polonik, Chair, Department of Statistics, UC Davis
Internships in Statistics

I. Concurrent Sessions	11:00-11:55

California Suite

(11:00-11:25) Bruce Leistikow, MD, MS, Associate Adjunct Professor of Medicine, School of Medicine, Department of Public Health Sciences, UC Davis
Can sentinel health events as exposure markers reduce misclassification, selection & detection bias and detect a 21st century 2 billion death epidemiologic booboo

Estimates that smoking will cause about 40 million US and 1 billion global deaths this century are underestimates due to selection, misclassification, & detection biases and consequent contrasts between unrepresentative light "smokers" and "nonsmokers" who smoked. New time-series and demographic research reports link smoking to the majority of male all sites cancer death rates in several groups and to most gender disparities in all cause mortality.

This talk will overview smoke exposures and effects and the origins, propagation, and potential work-arounds for those biases and their possible effects on estimates of mortality from smoking, obesity, and other exposures. The audience will participate in assessing 1) which were the most important missteps in this possible 21st century 2 billion death epidemiologic booboo; and 2) the utility of sentinel health event/chronic disease relationships in epidemiology.

(11:30 -11:55) Daniel Tancredi, PhD**,** Assistant Professor of Pediatrics, School of Medicine and Center for Healthcare Policy and Research, UC Davis
Bias-variance tradeoff for risk factor estimation when prevalent cases are excluded from epidemiological analysis of dementia incidence in standard two-timepoint design

The two-timepoint design is commonly used in epidemiological studies of dementia. A standard analysis strategy for assessing the effect of risk factors is to exclude from the follow-up incidence analysis all of the prevalent cases identified at study baseline. The exclusion of such cases is based on well-founded concerns about bias, but the loss of statistical information is substantial. Theoretical considerations and simulation results will be used to assess this bias-variance tradeoff in the setting of parametric survival analysis models.

Delta Suite

(11:00-11:25) Min Li, PhD**,** Assistant Professor, College of Business Administration, CSUS
Co-authors: Yan Yu, Keming Yu, and Hua Wang
Semiparametric Estimation for a Class of Time-Inhomogeneous Diffusion Processes

We develop two likelihood-based approaches to semiparametrically estimate a class of time-inhomogeneous diffusion processes: log penalized splines (P-splines) and the local log-linear method. Positive volatility is naturally embedded and this positivity is not guaranteed in most existing diffusion models. We investigate different smoothing parameter selections. Separate bandwidths are used for drift and volatility estimation. In the log P-splines approach, different smoothness for different time varying coefficients is feasible by assigning different penalty parameters. We also provide theorems for both approaches and report statistical inference results. Finally, we present a case study using the weekly three-month Treasury bill data from 1954 to 2004. We find that the log P-splines approach seems to capture the volatility dip in mid-1960s the best. We also present an application to calculate a financial market risk measure called Value at Risk (VaR) using statistical estimates from log P-splines.

(11:30 -11:55) Karla Van Meter, PhD, Researcher and Trainer, Van Meter Associates
Episcan as an R package: Improved power and flexibility in cluster detection

Episcan was developed to improve sensitivity and specificity of the spatial scan for cluster detection, particularly of irregular shapes. In simulation comparisons, the confocal ellipses of Episcan have performed as well or better than the most well known scan test. With Episcan, we report more information than just the single most likely cluster; instead a detected cluster is built by aggregating all clusters with high statistic and biologic significance to form the most significant cluster not limited by a prespecified form.

We are currently developing Episcan as an R package that will be epidemiologist-user-friendly. This includes tools for visualizing all significant clusters irrespective of overlap. Epidemiologists will be able to specify both a maximum p value and a minimum relative risk for the significant clusters that are included in a final map. We are interested in audience input on visualization tools that will be included, input format requirements or other features that would be useful.

Foothill Suite (Auburn)

(11:00-11:25) Enach Haga, PhD
Design of a statistical experiment to determine the theoretical limits of the English language

In 2000 I became interested in how much room the English language might have for expandability. The results of my investigation were published May 2000 in *Word Ways, The Journal of Recreational Linguistics*, whose editor, Ross Eckler, an expert in the field, gave advice along the way. At the time I focused on results rather than the method. Only later did I realize that what I had done was to design and carry out a statistical experiment. All I wanted was the answer, but that required lots of data. Just how I defined the problem, and decided on just what was needed for solution, and where to get it, is the subject of my presentation.

(11:30 -11:55) Lawrence C. Larsen**,** Staff Air Pollution Specialist, California Air Resources Board
How to consult with a statistician?

George Box defined statistics as "the art and science of making sense out of data." If this is true, then everyone who collects data needs to apply statistical thinking. Many who own data are insecure regarding statistics, so they decide to ask a statistician for help. The first sign of impending trouble may be a phone call that begins, "I just need a few minutes of your time." An experienced consumer of statistical consulting services knows better. If you lack experience consulting with statisticians, this session is for you. Two common issues will provide context: (1) I need to know how big a sample to collect. (2) I need to know whether my data are significant. Partial answers to questions like these will be offered: How much time and how many meetings will this take? What might the statistician ask me? What should I make sure I tell the statistician? What should I bring with me? Is it true that I should be afraid – very afraid?

Lunch Buffet	**Redwood Room**	**12:00- 12:55**

II. Concurrent Sessions	**1:00- 1:55**

California Suite

(1:00-1:25) Christina Drake, Ph.D, Associate Professor, Department of Statistics, UC Davis
Nonignorable nonresponse random subsamples of non-respondents and how to adjust for missing data

Longitudinal studies often experience loss to follow-up. The validity of inference depends on the missingness mechanism [Little J.A, and Rubin, D.B., 2002]. When the missing data mechanism depends on observed data only, estimation of means and/or regression coefficients requires adjustment. If the missingness mechanism depends on unobserved data, unbiased estimation requires further information. The information from random sub-samples of dropouts whose responses are obtained, can be used to model the data using selection or pattern mixture models [Allison, 1994], both of which are identifiable in this case. Multiple imputation [Rubin, D.B. 1987, Glynn et al., 1993] can lead to standard statistical analysis. Very large surveys can have more than 50% non-response. A naive approach using multiple imputation results in data sets with more than 50% imputed values. Another method is inclusion of the whole random subset of non-respondents with special follow-up and a random subset of respondents and other non-respondents for multiple imputation. We study mixture models for a binary outcome and propose a method for a common odds ratio from a mixture model. The methods are illustrated with the Project Talent data set.

(1:30 -1:55) Chris Barker, Ph.D, Postdoctoral Scholar, Center for Vectorborne Diseases, UC Davis
Hierarchical Bayesian modeling for vectorborne diseases

Hierarchical Bayesian (HB) models have become increasingly popular for modeling complicated ecological systems, such as those of vectorborne diseases. They offer several advantages, from modeling latent processes and autocorrelation to incorporation of prior knowledge for probabilistic inferences and predictions. Examples of HB models will be presented from two of our recent studies on mosquitoborne viruses in California. The first study used modeled long-term trends in wild bird abundance to determine whether populations declined following the recent arrival of West Nile virus. The second identified temporal linkages between mosquito abundance and virus transmission after adjustment for potential confounders. The relevance of HB models for analyzing retrospective surveillance datasets and accommodating their inherent complexity will be discussed.

Delta Suite

(1:00-1:25) Lois Lowe, PhD, Independent Consultant
Characteristics of California's Male Civil Addicts VS Male Felons, 1998 to 2008

In 1962, the California Department of Corrections (now Department of Corrections and Rehabilitation – CDCR) implemented the Civil Addict Treatment Program at the California Rehabilitation Center, Norco, for adults. Courts may sentence substance abusing offenders to the Civil Addict Program in lieu of a standard felony commitment. In the early years, an active treatment program existed. Subsequently treatment services declined. In 1995, CDCR decided that a more intensive program was required and augmented services. In 2008, CDCR issued a Request for Proposal for planning, developing & conducting a treatment program specifically for male civil addicts. Bidders were asked to describe the potential treatment population. Public sector data were used to construct behavioral and demographic statistics.

(1:30-1:55) Trac H. Pham, Research Program Specialist II, California Highway Patrol
Survey of California Highway Patrol Headquarters Employee Commuting Patterns, Attitudes and Behavior

This presentation reports on a recent statistical survey of the California Highway Patrol Headquarters employee commuting patterns, attitudes and behavior in preparation for planning their upcoming relocation to a new headquarters. It describes the survey design, data collection and statistical analysis of the findings. The coverage will include the online survey methodology, questionnaire design, basic statistical analyses of commuting patterns and behavior, and discuss resulting policy options to improve employee commuting and the driving environment. It will also deal with practical aspects of attitudinal surveying, including remedial measures to cope with low survey response rates, non-responses and missing data. As such, the presentation should help researchers, policy makers, and marketing managers on the use of statistical surveys to examine employee or consumer attitudes, needs and concerns.

Foothill Suite (Auburn)

(1:00-1:25) Ci-Ren Jiang, PhD Candidate, Department of Statistics, UC Davis
SSA 2009 Student Award Candidate
Covariate Adjusted Functional Principal Components Analysis for Longitudinal Data

Classical multivariate principal component analysis has been extended to functional data and termed Functional Principal Component Analysis (FPCA). Most existing FPCA approaches do not accommodate covariate information, and it is the goal of this paper to develop two methods that do. In the first approach, both the mean and covariance functions depend on the covariate Z and time scale t, while in the second approach only the mean function depends on the covariate Z. Both new approaches accommodate additional measurement errors and functional data sampled at regular time grids as well as sparse longitudinal data sampled at irregular time grids. The first approach to fully adjust both the mean and covariance functions adapts more to the data but is computationally more intensive than the approach to adjust the covariate effects on the mean function only. We develop general asymptotic theory for both approaches and compare their performance numerically through simulation studies and a data set.

(1:30 -1:55) Kehui Chen, PhD Student, Department of Statistics, UC Davis
SSA 2009 Student Award Candidate
Quantile Regression Conditioning on Functional Covariates with Application in Growth Curve Data

To investigate a future growth level, given currently observed growth path, estimating the conditional quantiles when covariates take values in functional space is a natural approach. In my presentation, I will briefly review previous work on estimating conditional quantiles when covariates are vectors. I will also present some preliminary ideas for our new proposed approach for the extension to the case of functional covariates, followed by the data analysis results for the Berkeley Growth Study data.

Orchard Suite (2nd floor)

(1:00-1:55) Paul T. Melevin, PhD, Research Program Specialist II, and
Susan Ayres, M.A., Research Program Specialists I, Survey and Applied Research Section,
Audit and Evaluation Division, State of CA Employment Development Department
Automating Dillman's Total Design Methodology for Mail Questionnaires

In 1978, Donald A. Dillman published the first edition of his book entitled, *Mail and Telephone Surveys: The Total Design Method*. In it, he outlined a methodology for inducing response to mail and telephone surveys that required an adherence to methods of personalization. In his second edition (Dillman, 2000:165) he indicated that the use of the various personalization elements produce a collective five to eight percentage point increase in response. At that time, he could not identify the degree to which each individual element contributed to this overall effect. In December 2007, we were confronted with trimming printing and manpower costs by eliminating several of these personalization elements from our mail survey efforts. We worked with our Publication and Mail Operations to partially automate the mainly manual mailing of our mail questionnaires. Because we were concerned with a lowered response rate, we conducted a comparative analysis of the response rates obtained using a personalized mailing conducted in the Spring of 2007 with the automated methods used in a similar period in 2008. While we were concerned that these changes might cause a decrease in response, the actual rate of response increased slightly in 2008. Further, the failure to find any decrease in response suggests that the automation methods can be used without harm to response. Additionally, since we charted daily returns, we can show that questionnaires were returned earlier using the automation.

Afternoon Plenary Session I	**Redwood Room**	**2:00- 2:45**

Featured Speaker
Jane-Ling Wang, Professor, Department of Statistics, University of California, Davis
Joint Modeling of Longitudinal and Survival Data

Refreshments	**Redwood Room**	**2:45- 3:00**

III. Concurrent Sessions	**3:00- 3:25**

California Suite
(3:00-3:25) Rani Celia Iaac, MBA, Economist and Researcher, State Library, California Research Bureau
Forecasting State Foreclosures

Rani has been publishing a quarterly California Foreclosure Watch (the latest covers Q4 2008 data) for the Assembly Banking & Finance Committee. She will discuss modeling efforts to incorporate the assumption that job losses must end and home prices must stabilize for the cycle to bottom out. The methods and data used in putting together the forecasts and the rationale behind them may be of interest to SSA members. She will touch upon the longer term ramifications of the crisis on state government and the private sector.

Delta Suite
(3:00-3:25) Gary He, PhD and David Rocha, BA, California Diabetes Program
Disparities in Prediabetes: Prevalence, Awareness, and Behavior

This is a cross-sectional study using the National Health and Nutrition Examination Survey (NHANES) 2005-2006 to examine disparities in prevalence of prediabetes across socio-demographic groups, and awareness of prediabetes and prevention-oriented behaviors in people with prediabetes. We explored associations between socio-demographic characteristics and prediabetes, classified as having impaired fasting glucose (IFG) or impaired glucose tolerance (IGT). Among the subset with prediabetes, we identified characteristics independently associated with awareness of prediabetes (those told by a clinician they had prediabetes or are at risk), and diabetes prevention-oriented behavior. We examined whether awareness of prediabetes was associated with behavior change efforts. Of 1,887 respondents age ≥18 who had both IFG and IGT tests, 558 had prediabetes, representing 29% of the U.S. adult population. In addition to age and BMI, Asian, males, speaking a language other than English at home and with lower educational attainment were associated with prediabetes. Among those with prediabetes, only 20.2% were aware of their heightened risk of diabetes. Among those with prediabetes, 51.4% reported attempting weight control, 47.9% increasing physical activity, and 53.7% lowering fat/calories in diet. Male gender (AOR=0.50) and lower educational attainment (AOR=0.53) were independently associated with inaction regarding weight control, physical activity and diet. Results suggest that the scope of socio-demographic disparities is compounded across individual steps in the prevalence>awareness>action pathway. Strategies to reduce incidence of prediabetes should combine glucose screening, case identification and counseling with multi-level interventions that can influence population-wide health behavior through alternate pathways.

Afternoon Plenary Session II	**Redwood Room**	**3:30-4:15**

Featured Speaker
Andrew H. Karp, Sierra Information Services
Unleashing the Power of Statistical Graphics in SAS 9.2 Software

Networking/Socializing	**Redwood Room**	**4:15-4:30**