



**SACRAMENTO  
STATISTICAL  
ASSOCIATION**

**PROGRAM  
Institute on Research and Statistics**

Wednesday, April 2, 2008  
California State University, Sacramento

A local chapter of the American Statistical Association

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

President: Gloria J. Robertson, 916-326-3726; Vice-President: Rahman Azari 530-752-7709; Secretary: Rafael Diaz 916-278-6221; Treasurer: Charles Chan 916-552-9694  
Past President: Christina Drake, 530-752-8170; ASA Representative: Linda Gage, 327-0103 x2549; Councilors: Tadese Alemu 916-322-4086, Shannon Conroy, 916-208-3161, Rose Cendak, Sophie Flaherty 916-262-2234, Farzaneh Tabnak 916-324-0128

All events will take place at the California State University Union

<b>Registration and Continental Breakfast</b>	<b>Redwood Room</b>	<b>8:00-9:00</b>
-----------------------------------------------	---------------------	------------------

<b>Morning Plenary Sessions</b>	<b>Redwood Room</b>	<b>9:00-11:30</b>
---------------------------------	---------------------	-------------------

**9:00-9:15**

**Welcoming Remarks**

Gloria J. Robertson, President, Sacramento Statistical Association  
Jill M. Trainer, Dean, School of Natural Sciences and Mathematics, CSUS

**9:15-10:00**

**Featured Speaker**

Howard Roth, Chief Economist, California Department of Finance  
*Will the California Economy Go into Recession in 2008?*

**10:00-10:45**

**Featured Speaker**

Jane-Ling Wang, Professor, Department of Statistics, and Director, Statistics Laboratory,  
University of California, Davis  
*Death and Success on Mount Everest: Is there a Kareem Abdul-Jabbar effect?*

**10: 45 – 11:30**

**Featured Speaker**

Karla C. Van Meter, PhD, Department of Public Health Sciences, University of California, Davis  
*Geographic Distribution of Autism in California: A Retrospective Birth Cohort Analysis*

<b>I. Concurrent Sessions</b>	<b>11:35-12:00</b>
-------------------------------	--------------------

**California Suite  
(11:35-Noon)**

Danh V. Nguyen, Assistant Professor, Division of Biostatistics, Department of Public Health Sciences,  
University of California, Davis  
*Covariate Adjusted Correlation with Application to Female Fragile X Premutation Carriers*

Motivated by molecular data on female premutation carriers of the fragile X mental retardation 1 (*FMR1*) gene, we present a new method of covariate adjusted correlation analysis to examine the association of messenger RNA (mRNA) and number of CGG repeat expansion in the *FMR1* gene. The association between the molecular variables in female carriers needs to adjust for activation ratio (ActRatio), a measure which accounts for the protective effects of one normal X chromosome in female carriers. However, there are inherent uncertainties in the exact effects of ActRatio on the molecular measures of interest. In order to account for these uncertainties, we develop a flexible adjustment that accommodates both additive and multiplicative affects of ActRatio nonparametrically. Application to *FMR1* premutation data on 165 female carriers indicates that the association between mRNA and CGG repeat after adjusting for ActRatio is stronger and is more in-line with association seen in male carriers. Finally, the results provide independent support for a specific jointly additive and multiplicative adjustment form for ActRatio previously proposed in the fragile X premutation literature.

**Delta Suite  
(11:35-Noon)**

Rani Celia Issac, State Library, California Research Bureau  
*Forecasting State Housing Demand*

As foreclosure rates rise, home sales slump, and new building continues to slow, economists are trying to estimate how many potential homebuyers there are in order to estimate the number of years the turnaround in the housing market will take. Projections vary depending upon the population

data used - either the State of California's Department of Finance Demographic Research Unit's or the Census Bureau's. Rani recently published a report that explores how differences in the number of people, number of homes, vacancy rates, and household size impact the forecast. The report uses historical starts and permit data coupled with demographic trend analysis like migration and growth in the labor force.

Housing starts are among the most complex regional variables to forecast. There are two reasons for this – a lack of data and the nature of the industry. Only permit data is available and permits don't always become starts. With multifamily housing there is often a significant lag and sometimes in a downturn the lags can stretch out over years. These permit data usually only cover a portion of each state (what the Census Bureau refers to as permit-issuing places). Thus, historical data on housing starts by state must be estimated from the limited coverage permit data. This is an inexact process that can only be verified in years when the Census of Housing is conducted (1980, 1990, 2000, etc.). The second problem is that housing starts are extremely volatile, responding rapidly to interest rate changes, credit availability, changes in vacancy rates, weather, job growth or decline, and other factors. The annual changes that occur in starts are added to the level of total housing but must be netted of houses that are removed or demolished. Forecasting the level of housing starts is analogous to forecasting the change in employment or the change in Gross Domestic Product rather than their levels. In effect, housing starts measure a change and are a flow variable while total housing stock is a level. Forecasts for second homes and vacancy rates are factored into total stock estimates to 2010.

---

### **Foothill Suite (Auburn)**

**(11:35-Noon)**

Fushing Hsieh<sup>1</sup>, Li Zhu<sup>1</sup>, David I. Shapiro-Ilan<sup>2</sup>, James F. Campbell<sup>3</sup> and Edwin E. Lewis<sup>3</sup>

<sup>1</sup>Department of Statistics, University of California, Davis <sup>2</sup>USDA-ARS, GMPCRC, Manhattan, KS ,

<sup>3</sup>USDA-ARS, Southeast Fruit and Tree Nut Research Lab, Byron, GA, <sup>4</sup>Department of Entomology and Department of Nematology, University of California, Davis

*State-space based mass event-history model -- many decision-making agents with one target*

A dynamic decision-making system that includes a mass of indistinguishable agents could manifest impressive heterogeneity, which is postulated to result from microscopic behavioral tactics employed by almost all involved agents. A State-Space Based (SSB) mass event-history model is developed here to explore the potential existence of such microscopic behaviors. By imposing a latent internal state-space variable into the system, each individual's event-history can be decomposed into a common state duration and an individual-specific time to action. Thereby parametric statistical inferences are derived under the interval censorship and conditional independence assumptions. Identifiability and computation-related issues will also be addressed. From the dynamic perspectives of system-wise heterogeneity, this SSB mass event-history model shows essential difference from a random effects model via the Principle Component Analysis (PCA) based on numerical simulation. The proposed approach is demonstrated in a case study for mass invasion of parasitic nematodes into host larvae. The analysis results not only showed coherence in the context of the biology of the nematode as a parasite, but also provided new quantitative interpretations.

<b>Lunch Buffet</b>	<b>Redwood Room</b>	<b>12:00-1:00</b>
---------------------	---------------------	-------------------

<b>II. Concurrent Sessions</b>	<b>1:00-2:00</b>
--------------------------------	------------------

### **California Suite**

**(1:00-1:30)**

Farzaneh Tabnak<sup>1</sup>, Hans Müller<sup>2</sup>, Jane-Ling Wang<sup>2</sup> and Weihong Zhang<sup>1</sup>

<sup>1</sup>Cancer Detection Section, California Department of Public Health

<sup>2</sup>Department of Statistics, University of California, Davis

*Analysis of Time Characteristics of Breast Cancer Prevention in a Cohort of Underserved California Women*

Timely diagnosis and treatment following breast cancer screening will not only lead to reduction in mortality, but also morbidity that is related to anxiety from being informed of abnormal results. This research examines the influence of screening history and clinical and demographic characteristics of participants on the timeline of diagnosis and treatment after abnormal findings in the breast cancer screening cycle. More than 900,000 mammography screening cycle data from the California Department of Public Health, *Cancer Detection Programs: Every Woman Counts* (CDP: EWC) were analyzed. Waiting times from screening to diagnosis and from diagnosis to treatment were used as outcomes and were related to characteristics of women through the Cox proportional hazards regression model. In addition, logistic regression models were employed to estimate the odds that screened women of certain characteristics were being scheduled for final diagnosis. Our analyses suggested that the scheduling of final diagnosis, time from screening to final diagnosis and time from final diagnosis to treatment are affected by participants' clinical and demographic characteristics as well as the screening history.

**(1:30 -2:00)**

Coskun Cetin, Assist. Professor, Department of Mathematics and Statistics, California State University, Sacramento, and Rafiqul Bhuyan, Associate Professor, California State University, San Bernardino

*Optimal management of initial public offerings (IPOs) in a dynamic information-theoretic setting*

This paper is a contribution to the finance literature to explain the underpricing phenomenon and the long run under performance puzzles associated with the initial public offering (IPO) of the firms in the presence of asymmetric information. In our model, the owner-manager of the firm signals to the investors by means of information production and reduced IPO price initially and then through the dividend amounts subsequently. Using

Bayesian approach and linear programming methods, the model predicts the optimal behavior of the owners for different quality levels of the firms in a dynamic setting and distinguishes the quality type of the firms eventually.

---

**Delta Suite****(1:00-1:30)**

Alice van Ommeren and Willard Hom, Research & Planning, California Community Colleges  
*Clustering California Community Colleges on Environmental Factors Affecting Transfer*

Transfers from California's community colleges to four-year institutions continue to be an important measure of performance for the community colleges, as well as for the system of higher education. This presentation discusses several clustering methods for, and presents some of the results of, the most current community college transfer rates. The transfer rate methodology involves tracking for six years several cohorts of students who show behavioral intent to transfer. The rates account for differences among the colleges by adjusting for environmental factors affecting transfer and grouping college that are similar.

The uncontrollable variables attempt to account for the environmental differences among the colleges. We used hierarchical regression to identify the variables that predicted transfer performance: (a) percentage of students 25 years or older at a community college and (b) an index representing the bachelor degree attainment of the college's service area. The environmental factors were used to adjust the transfer rates and identify clusters of colleges with similar characteristics related to transfer, allowing colleges to evaluate their transfer performance among peers or "like" colleges. The results of several different clustering methods will be presented.

**(1:30- 2:00)**

Ori Stitelman, Alan Hubbard and Michelle Pearl, Department of Biostatistics, University of California, Berkeley  
*Applying Nonparametric Locally Efficient Estimation of the Treatment Specific Survival Distribution To Time To Pregnancy Studies*

Parametric models, including the proportional odds model and beta geometric model, are typically used to evaluate the effect of a treatment on fertility in time to pregnancy studies. However, these models rely heavily on parametric assumptions and will be inconsistent in many situations. This is particularly true in the case where censoring time is informative. The use of nonparametric locally efficient estimation of the treatment specific survival distribution as proposed by Hubbard et al. in 2001 provides an alternative methodology which results in estimates which are asymptotically efficient if the working model contains the truth and are consistent otherwise. Moreover, these estimators adjust for confounding and allow for dependent censoring.

The original methods discussed above will be applied to a study designed to evaluate the effects of second hand tobacco smoke (SHS) on fertility. In addition, the nonparametric locally efficient estimation methodology will be proposed as an alternative method to use in time to pregnancy studies which makes up for some of the shortcomings of the other two models. Furthermore, the benefits and deficiencies of each methodology will be explored.

---

**Foothill Suite (Auburn)****(1:00-1:30)**

Shuang Wu, Department of Statistics, University of California, Davis  
*Functional Data Analysis for Bid Arrival times in eBay Auctions*

Consider a situation where one observes a relatively small number of irregularly spaced repeated events for a number of subjects. An example is the bid arrival times for online eBay auctions, which follow certain typical patterns but where many of the auctions draw only few bids. We model bid arrivals by postulating a random rate function as realization of an underlying stochastic process for each auction. Focusing on the random density function that reflects the shape of the rate function, a functional data analysis approach is developed by estimating mean and covariance structure of these random density functions, followed by the estimation of eigenfunctions and functional principal component scores to obtain fitted density functions for individual auctions. Theoretical results include asymptotic consistency for the components of the model. The method is illustrated with a study of bid arrival times for 156 Palm M515 PDA auctions from eBay.

**Foothill Suite (Auburn)****(1:30-2:00)**

Meng Mao, Department of Statistics, University of California, Davis  
*Semi-parametric Efficient Estimation in the Proportional Odds Cure Models*

We present a mixture cure model with the survival time of the "uncured" group following the proportional odds property. The EM algorithm is used to locate the nonparametric maximum likelihood estimator (NPML). Unlike the proportional hazard mixture model, there is difficulty in the M step triggered by the nonparametric component, so we propose a hybrid algorithm by applying the MM algorithm in the beginning and then switch to the Newton-Raphson algorithm. This hybrid algorithm provides both computational stability and efficiency as demonstrated by simulation studies. Consistency and asymptotic efficiency of the nonparametric maximum likelihood estimator are derived. A case study of leukemia data is conducted to illustrate the proposed procedure, which is also compared to an existing procedure in a simulation study.

**Featured Speaker**

George G. Roussas, Professor, Department of Statistics, University of California, Davis  
*Why we do what we do in Mathematical Statistics: A rigorous justification of some statistical practices*

**Afternoon Refreshments**

2:45-3:00

**III. Concurrent Sessions**

3:00-3:30

**California Suite**

(3:00-3:30)

Julie Yee<sup>1</sup>, Bill Perry<sup>1</sup>, Gerry McChesney<sup>2</sup><sup>1</sup>U.S. Geological Survey, Western Ecological Research Center,<sup>2</sup>U.S. Fish and Wildlife Service, San Francisco Bay National Wildlife Refuge Complex*The Common Murres of Devil's Slide Rock: How a seabird colony recovered from a devastating oil spill*

In 1986, the barge *Apex Houston* spilled oil into the ocean killing approximately 9000 seabirds including 6300 common murres (*Uria aalge*). Just south of the San Francisco Bay, a colony site for breeding common murres at Devil's Slide Rock was abandoned. Ten years later, the U.S. Fish and Wildlife Service and cooperators began implementing a plan to restore the extirpated colony. Using social attraction techniques, common murres were lured back to the rock and a colony was reestablished. In an effort to better understand the factors related to the success of colony restoration, these murres were intensively observed and data on their breeding attempts, selection of nest sites, and breeding outcomes were recorded. GIS analysis indicates that the nesting success may be spatially related, and logistic regression and regression tree models suggest that social factors such as proximity to nearest neighbors and density of surrounding breeding pairs may be important to the success of breeding attempts.

**Delta Suite**

(3:00-3:30)

Gary He, PhD and David Rocha, BA, California Diabetes Program

*Models of Diabetes Risks Assessment using Self-Reported Measures*

Objectives: National data suggest that prevalence of prediabetes may be greatly underestimated based on self-report of diagnosed prediabetes from state-level diabetes surveillance data sources, most commonly the Behavioral Risk Factor Surveillance System (BRFSS). The goal of this analysis is to develop models to predict diabetes risks including prediabetes and undiagnosed diabetes using self-reported measures (SRM) available from BRFSS.

**Foothill Suite (Auburn)**

(3:00-3:30)

Ruixiao Lu, PhD, Department of Data Analysis and Algorithm, Affymetrix, Inc.

*Using Biological Information to Increase the Sensitivity of Tests for Differential Gene Expression in Microarray*

Identification of differential expression of transcripts, proteins, and other biological molecules is complicated by a number of factors. First, the very comprehensiveness of high-throughput methods makes it difficult to separate real effects from statistical accidents. Second, there are differences in timing of transcriptional cascades and other molecular pathways between individual samples-this is especially true for human data. We have developed a new method for dealing with this problem by utilizing biological information in the analysis, after careful low-level preprocessing of the raw signals and statistical analysis of the significance of changes in single probes or genes. The first step is to select one or more pathways or gene groups and identify the probes/probe sets on an array that correspond to the pathway or gene group. Second, for each probe/probe set, we conduct the usual statistical analysis such as a t-test for the difference between treatment and control groups. We then test for whether in the aggregate this shows up-regulation, down-regulation, or neither. The significance of the test could be accessed by sampling genes or sampling arrays, and we show the pros and cons for both of them. This new method can also serve as a powerful tool for evaluating the reproducibility of a study. We illustrate the value of the method with results from dose-response study of low dose radiation exposure data from human expression arrays. An application study and the full theoretical methodology will be described in details.

**Featured Speaker**

Jerome Braun, PhD, Statistics Laboratory, University of California, Davis  
*Using the SAS package for modeling dependent data structures.*

**Networking/Socializing**

4:00