

of the most sensitive group. These findings suggest that participants in the negative condition may have utilized the aid less than those in the neutral condition. This indicates that the potential emotional experiences of an operator should be considered when determining when and where to integrate automated decision aids.

HOW THE FRAMING OF METAPHORS AFFECTS PUBLIC POLICY TOWARD THE MENTALLY ILL. Victoria E. Bennett & Della N. Gibson, Department of Psychology, University of Mary Washington, Fredericksburg, VA 22401. Metaphors have been shown to influence people's decisions about campaigns, ads, and politics. Specifically, studies have looked at how metaphors have influenced attitudes toward crime and obesity. We examined the effect of blaming and victimization metaphors on public policy attitudes toward the mentally ill. It was hypothesized that victimization metaphors, as compared to blaming metaphors, would lead to increased support of outpatient treatment, increased support of government funding, and increased support of rights and opportunities for the mentally ill. The results showed no significant differences between the three conditions. Additionally, metaphors may not influence people with strongly held beliefs toward an issue.

A CONCEPTUAL MODEL FOR BYSTANDER INTERVENTION: EXPANDING PROSOCIAL BYSTANDER BEHAVIOR. M. Sihalath, B. Wright, P. Pierucci & R. Cobb-Ozanne, Center for Applied Behavior Systems, Department of Psychology, Virginia Tech, Blacksburg VA 24060. The frequency of bullying and interpersonal violence remains high throughout our culture. In these situations where intervention is needed, bystanders play a vital role in potentially reducing interpersonal violence. Bystanders in emergency or violent situations choose whether or not to intervene for the prevention of harm and/or injury. This extends the bystander intervention model from the prevention of interpersonal harm to the promotion of prosocial behavior. Participants were given green wristbands to recognize desirable prosocial behavior and were measured on the four factors from the Latané and Darley model (i.e., "notice", "interpret", "responsibility", and "skills"). In regards to predicting prosocial recognition, the results suggest "responsibility" and ability to "interpret" were significant predictors. In regards to predicting intention to intervene, the data suggests "notice" to be a significant predictor.

Statistics

STATISTICAL PREDICTION FOR VIRGINIA LYME DISEASE EMERGENCE BASED ON SPATIAL-TEMPORAL COUNTS DATA. Yuanyuan Duan¹, Jie Li¹, Yili Hong¹, Korine N. Kolivras², James B. Campbell², Stephen P. Prisley³ & David N. Gaines⁴, ¹Department of Statistics, Virginia Tech, Blacksburg, VA 24061², Department of Geography, Virginia Tech, Blacksburg, VA 24061, ³Department of Forest Resources and Environmental Conservation, Virginia Tech, Blacksburg, VA 24061 & ⁴Virginia Department of Health, Richmond, Virginia 23219. The emergence of infectious diseases over the past several decades has highlighted the need to better understand and prepare for epidemics as endemic infectious diseases. These diseases are usually expanding their geographic range and are recorded over multiple time periods, making

the analysis and prediction more complicated. This study is based on areal (census tract level) counts data of Lyme disease cases in Virginia from 2003 to 2010. Our visualization of incident rate using kernel smoothing showed the spatial-temporal emergence pattern of Lyme disease. We built a spatial Poisson regression model with random effect to incorporate spatial-temporal correlations. The random effects were modeled by Conditional Auto-Regressive (CAR) model. We used Markov chain Monte Carlo (MCMC) algorithm to fit the model in a Bayesian framework. Using this model, we analyzed 5 year sum counts from 2006 to 2010 to identify significant landscape/demographic variables for different eco-regions. We also analyzed yearly counts data and predict 2011's count based on previous years. This study was funded by NSF-BCS-GSS Grant.

PRODUCT COMPONENT GENEALOGY MODELING AND WARRANTY RETURN PREDICTION. Caleb King¹, Yili Hong¹ & William Q. Meeker²,¹Department of Statistics, Virginia Tech, Blacksburg, VA 24061 & ²Department of Statistics, Iowa State University, Ames IA 50011. Many industrial products consist of multiple components that are necessary for maintaining their integrity. There is already an abundance of literature on modeling the lifetime of such components through competing risks models. However, none of these models take into consideration the fact that many of these components are part of a specific generation of components and that each of these components may go through several generations throughout a period of observation. We believe that by incorporating this extra information, which is readily available in most product warranty databases, better accuracy can be achieved in predicting time to failure thus yielding more reliable warranty return information. We present methods for estimating parameters for this new generational model as well as provide a comparison with existing methods through the use of simulation. Our results indicate that the generational model outperforms the existing methods in predicting warranty returns.

DEGRADATION DATA ANALYSIS USING NONLINEAR MODELS VIA SHAPE-RESTRICTED SPLINES. Zhibing Xu & Yili Hong, Department of Statistics, Virginia Tech, Blacksburg, VA 24061. The traditional reliability analysis is based on the failure-to-time data. However, many products are designed with high reliability and low risk for failures. It is hard to collect enough failure-to-time data to give a precise estimation for the highly reliable products in a limit time, even under accelerated conditions. Comparing to the traditional life time data, degradation data not only can provide enough information in a short time, but also give a better estimation and prediction of the life time of products. With the improvement of technology, life-affecting environmental variables are recorded as well as the degradation measure over time. For example, the large amount of ultraviolet exposure will accelerate the degradation of the polymer coating. Thus, it is important to incorporate the environmental variables, also called dynamic covariates, into the degradation path model. Existing research suggests a linear mixed effect model with splines to combine the dynamic covariates. With unspecific effect function for the dynamic covariates, shape restricted splines method is used. However, the linear model may not widely used in real applications, because most of the degradation data follow a nonlinear trend, especially in the pharmacokinetic research. In this paper, we propose a nonlinear mixed

effect model combining with dynamic covariates to model the degradation data. Shape restricted splines are used in the proposed model and a modified alternating algorithm is developed. The performance of the algorithm is evaluated by simulations. An outdoor weathering dataset is used for illustration of the proposed method.

DECISION TREES WITH HIGH DIMENSIONAL DATA AND EXPLICIT COVARIATE SELECTION. Lucas Roberts, Department of Statistics, Virginia Tech, Blacksburg, VA 24061. In this talk, we survey the decision tree literature from both a model based and a heuristic perspective. Primarily, these methods have been successfully deployed in cases where prediction is of interest. When we are mainly interested in inference, which in this context means identifying the most useful covariates, these methods can exhibit poor behavior. These inferential problems are exacerbated in high dimensional spaces. After reviewing some general covariate filtering approaches, we present our method, which does covariate filtering and decision tree search jointly. We state necessary and sufficient conditions ensuring consistency of the tree classifier and we show the benefits of our approach through simulated and real examples. We also give a rule for ranking and selecting covariates using output from our model. We compare our covariate ranking and selection procedure to current common approaches.

CLUSTER-BASED PROFILE MONITORING IN PHASE I ANALYSIS. Yajuan Chen, Jeffery B. Birch & William Woodall, Department of Statistics, Virginia Polytechnic Inst. & State Univ., Blacksburg VA 24061. An innovative profile monitoring methodology is introduced for Phase I analysis. The proposed technique, referred to as cluster-based profile monitoring, incorporates a cluster analysis phase to aid in determining the possible existence of “profiles” in the historical data set (HDS) resulting from an out-of-control process. To cluster the profiles, the proposed method first replaces the data from each sampled unit with an estimated profile, using some appropriate regression method, and clusters the profiles based on their estimated parameters. This cluster phase then yields an initial main cluster which contains at least half the profiles. The initial estimated population average (PA) profile is obtained by fitting a linear mixed model to those profiles in the main cluster. Profiles, determined using the Hotelling’s T² statistic, that are not contained in the initial main cluster are iteratively added to the main cluster and the mixed model is used to update the estimated parameters for the PA profile. Those profiles contained in the final main cluster are considered as resulting from the in-control process while those not included are considered as resulting from an out-of-control process. A simulated example and Monte Carlo results demonstrate the performance advantage of this proposed method over a current non-cluster-based method with respect to more accurate estimates of the PA parameters.

ESTIMATION AND EVALUATION OF OPTIMAL THRESHOLDS FOR THE BELIEVE THE POSITIVE SEQUENTIAL TESTING STRATEGY. Amber R. Wilk & Donna K. McClish, Department of Biostatistics, Virginia Commonwealth University, Richmond VA 23219. Many continuous medical tests rely on a threshold for diagnosis. A sequential testing strategy, called Believe the Positive, classifies a patient positive if either the first test is greater than a threshold θ_1 or negative on the

first test and greater than θ_2 on the second test. A threshold pair $\theta = (\theta_1, \theta_2)$ was defined as optimal if it maximized $GYI = \text{Sensitivity (Se)} + r^*(\text{Specificity (Sp)} - 1)$. Of interest is to determine if the optimal threshold pair estimates are “good” when calculated from a sample. Formulas were derived to estimate θ^* assuming tests follow a binormal distribution, using the Newton-Raphson algorithm. A simulation study was performed assessing bias, root mean square error (RMSE), percentage of under/over estimation of Se/Sp, and coverage of simultaneous confidence intervals (SCI)/ellipses for sets of population parameters and sample sizes. Bootstrapping was used to estimate the variance of each optimal threshold pair. When the area under the curve (AUC) between the two tests is equal, the median RMSE and median bias in Se and p appears to decrease. The coverage of the SCI and confidence ellipses improves as sample size, AUC, and the ratio of standard deviations of each group increases. This also holds true for percentage of overestimation of Se and Sp except when the allocation between groups favors those without disease. This method is an improvement over an empirical estimate.

LIKELIHOOD BASED FORECASTING OF LOCALLY CHANGING VARIANCE STRUCTURED TIME SERIES DATA. Rajan Lamichhane & Norou Diawara, Dept. of Mathematics and Statistics, Old Dominion University, Norfolk VA 23529. In this talk, we assume several non-linear time series models with additive noise components, and the model fitting is proposed in two stages. The first stage identifies the density ratio using all the clusters information, without specifying any prior knowledge of the underlying distribution function. In the second stage, we partition the time series into consecutive non-overlapping intervals of quasi stationary increments where the coefficients shift from one stable regression relationship to a different one using breakpoints detection algorithm. These breakpoints are estimated by minimizing the likelihood from the residuals. We develop a method based on finite mixture of Gaussian distributions to forecast this type of partitioned data. The Expectation-Maximization (EM) algorithm, with initial values obtained from the empirical estimates; give the estimates of mixture distribution. We propose likelihood based method to identify the most probable value from a pool of past values for the forecasting.

ESTIMATING EXPLAINED VARIATION FOR AN UNDERLYING LINEAR MODEL USING LOGISTIC REGRESSION. D. Sharma, Department of Mathematics and Statistics, James Madison University, Harrisonburg, VA 22807. The coefficient of determinant, also known as the R^2 statistic, is used to assess the extent of the strength of the relationship between a response and explanatory variables in a linear regression model. It measures the proportion of variation in the response variable explained by a set of independent variables. In many real life events, interest lies on modeling the relationship between a continuous response variable and a set of predictors. But in practice, the continuous dependent variable of interest may not be observable and is usually represented by its binary proxy. In such situations, logistic regression is a popular choice. There are many R^2 type statistics proposed to measure of explained variation for logistic regression. The pseudo R^2 measure (R_L^2) stands out because of its intuitive interpretation and independence on the proportion of success in the sample. It, however, severely underestimates the proportion of explained

variation (P^2) in the variable underlying the binary indicator of event occurrence. In this research we present a method for estimating the explained variation for the underlying linear model using the pseudo R^2 statistics obtained from a logistic regression analysis.

STRUCTURED FUNCTIONAL ADDITIVE REGRESSION IN REPRODUCING KERNEL HILBERT SPACES. Hongxiao Zhu¹, F. Yao² & H. H. Zhang³, ¹Department of Statistics, Virginia Tech, Blacksburg, VA 24061, ²Dept. of Statistics, University of Toronto, & ³University of Arizona. Functional additive models (FAMs) provide a flexible yet simple framework for regressions involving functional predictors. The utilization of data-driven basis in an additive rather than linear structure naturally extends the classical functional linear model. However, the critical issue of selecting nonlinear additive components has been less studied. In this work, we propose a new regularization framework for the structure estimation in the context of Reproducing Kernel Hilbert Spaces. The proposed approach takes advantage of the functional principal components which greatly facilitates the implementation and the theoretical analysis. The selection and estimation are achieved by penalized least squares using a penalty which encourages the sparse structure of the additive components. Theoretical properties such as the rate of convergence are investigated. The empirical performance is demonstrated through simulation studies and a real data application.

DICTIONARY LEARNING FOR GENETIC DATA IN RELATED INDIVIDUALS. Xiaowei Wu & H Zhu, Department of Statistics, Virginia Tech, Blacksburg, VA 24061. Dictionary learning has been widely used for sparse representation and data restoration in computer vision and image analysis. However, its application in statistical genetics has not been fully explored. We consider the problem of learning the sparsity structure of genetic data in related individuals. We propose a nonparametric Bayesian method based on a latent beta process factor analysis model. The method can be used in association study with related individuals for multiple purposes, for example, genotype imputation, estimation of allele frequency or kinship coefficients. Simulation study shows good performance and advantages of the proposed method in comparison with some existing methods. We also apply our method to the analysis of the GAW 18 data set.

EVALUATING THE HEALTH EFFECT(S) OF ENVIRONMENTAL CHEMICALS AND NUTRITION: CONSIDERATIONS FOR HIGHLY CORRELATED DATA. Chris Gennings, Virginia Commonwealth University, Richmond VA 23298. Biomonitoring of environmental chemicals in human tissues and fluids has shown that all people, not just those working in or living near major pollution sources, carry a “body burden” of synthetic chemicals in their blood, fat, mother’s milk, semen, urine and breath. Despite increasing public concern regarding the potential health effects of pervasive and persistent environmental exposures, there are few, if any, recommendations for life style changes to mitigate such effects. Our research focus is a holistic evaluation of the impact of environmental chemicals on health effects including metabolic endpoints (e.g., obesity, hyperlipidemia) and hepatotoxicity and the role nutrition may play in potentially mitigating the impact of chemical exposures. As such, we are not focused on single chemicals or nutrients – instead, we are focused

on their potential combined effects. Standard statistical methods suffer from multicollinearity effects due to the complex correlation structure among the components. We have developed novel statistical methods to determine empirical weights that identify subsets of components most associated with a response variable. The results are weighted indices of body burden and nutrients – i.e., simple to understand and visualize indices that may be easily combined with other behavioral data for additional analyses. Characterization of the approach will be described through simulation studies. The methods will be demonstrated using NHANES data.

Structural Biology, Biochemistry, and Biophysics

UBIQUITIN MODULATES TOLLIP'S LIPID-MEDIATED ENDOSOMAL MEMBRANE BINDING. S. Mitra, A. Traughber, S. Gomez, & D. G. S. Capelluto. Department of Biological Sciences, Virginia Tech, Blacksburg, VA 24061. Ubiquitylation is a highly controlled post-translational modification of proteins, in which proteins are conjugated either with monoubiquitin or polyubiquitin chains. Ubiquitin modifications on target proteins are recognized by ubiquitin-binding domains, which are found in several effector proteins. In this study, we describe for the first time how ubiquitin controls the function of the Toll-interacting protein (Tollip), which is an effector protein in the innate immune signalling pathway and an adaptor protein for endosomal trafficking. We have demonstrated that the central C2 domain of Tollip preferentially interacts with phosphoinositides. Remarkably, we have observed an ubiquitin dose-dependent inhibition of binding of Tollip to phosphoinositides and it does so specifically by blocking Tollip C2 domain-phosphoinositide interactions. This led us to discover that the Tollip C2 domain is a novel ubiquitin-binding domain. In addition, we have biophysically characterized the association of the Tollip CUE domain to ubiquitin and compared it with Tollip C2 domain-ubiquitin binding. We have also found that ubiquitin binding to dimeric Tollip CUE domain induces a drastic conformational change in the protein, leading to the formation of a heterodimeric Tollip CUE-ubiquitin complex. These data suggest that ubiquitin binding to the Tollip C2 and CUE domain and ubiquitin-mediated dissociation of CUE dimer reduces the affinity of Tollip protein for endosomal phosphoinositides, allowing Tollip's cytoplasmic sequestration for its cytosolic commitments.

MECHANISTIC STUDIES OF A FLAVIN DEPENDENT LYSINE MONOOXYGENASE FROM *NOCARDIA FARCINICA*. R.M. Robinson, P. Rodriguez, N. Keul, & P. Sobrado, Department of Biochemistry, Virginia Tech, Blacksburg VA 24061. The kinetic and chemical mechanism of the lysine monooxygenase Nocobactin G (NbtG) from *Nocardia farcinica* was investigated using both steady-state and rapid reaction kinetics. NbtG hydroxylates both L- and D-lysine, a unique characteristic among members of the *N*-hydroxylating monooxygenases (NMOs), which are typically selective for the L-stereoisomer of their respective substrate. NbtG is mostly uncoupled (~75%) and produces superoxide and hydrogen peroxide as oxidative by-products. O₂ consumption increases from 0.581 s⁻¹ in the absence of lysine to 1.09 s⁻¹ and 3.03 s⁻¹ when saturating concentrations of L- and D-